Chapter 13

Distance and continuity

Introduction

You already know how useful and important it is to be able to determine the distance between points in the plane (or space); here you will see whether it is possible to define a useful notion of distance between more abstract objects such as functions, or elements from a group. In this book (and continuing in Book F), we develop a theory that enables us to talk about the distance between elements from a diverse range of different types of object (such as groups and function spaces) by introducing the theory of *metric spaces*: sets that have been equipped with some suitable method for determining a 'distance' between any two elements.

Although (as you will see in the next chapter) the definition of a metric space is a broad one, it gives rise to a rich theory that generalises and extends many ideas from (real) analysis with which you will already be familiar. In particular, it enables us to talk about the convergence (or otherwise) of sequences of objects other than real numbers, such as functions or sets. This in turn provides a way to extend the concept of continuity far beyond that of functions on the real line, ultimately enabling us to confirm that problems such as the following have (unique) solutions:

> Is there a continuous function $f\colon [0,1] \to [0,1]$ such that $f(0) = 0$ and, for $0 < s < 1$, f satisfies the differential equation
>
> $$\frac{df}{ds} = \tfrac{1}{2}\cos f(s)?$$

We will develop a method for showing that problems like this have a solution in Book F, Chapter 23, *Completeness*. Note that Wolfram Alpha gives the solution of the ordinary differential equation to be $f(s) = 2\tan^{-1}(\tanh(s/4))$.

This first chapter has a different character from that of the subsequent chapters in this book, in that much of it revises concepts from real analysis with which you should already be familiar from your previous studies. It begins with a review of sequences and functions of real numbers, followed by a reminder of the definition of continuity for functions on the real numbers.

This raises a question: how might we extend the notion of continuity to functions between more general sets? For example, which functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ should we describe as 'continuous'? As a first step towards answering this question, in Section 3 we generalise the definition of continuity for functions from $\mathbb{R}$ to $\mathbb{R}$ to functions from $\mathbb{R}^2$ to $\mathbb{R}$. A crucial aspect of this process involves uncovering certain properties embedded in our idea of the 'distance between points'. In the next chapter, a further exploration of these concepts will lead to definitions of 'distance' and 'continuity' applicable not only to $\mathbb{R}^n$, but also in a much more general setting.

Chapter guide

Most of the material in Sections 1 and 2 should already be familiar to you, so it is likely that you will not need to spend that much time working through these sections. You might find it helpful to use the exercises to identify the parts that you need to concentrate on.

In Section 3 we start to look at a new situation and examine how our understanding of continuity for functions on the real line can be used to motivate a definition of continuity for functions mapping from the plane to the real line; this is our first tentative step towards metric spaces.

1 Functions and sequences

We start our investigation of continuity by clarifying some of the important terminology and notions on which the definition of continuity for functions from $\mathbb{R}$ to $\mathbb{R}$ is based.

As in previous books of M303, we assume that you are familiar with the usual arithmetic operations of addition, subtraction, multiplication and division of real numbers; the different sorts of interval on the real line; manipulating simple inequalities involving real numbers; and the trigonometric and exponential functions.

A brief summary of the concepts with which you should be familiar is given in the first part of the *Handbook*.

We also assume that you are already familiar with the basic ideas associated with sets and subsets, such as membership, union and intersection. Many of these ideas were used in Book B.

As a reminder, we list some frequently seen sets of numbers, most of which you have already encountered in Book A and elsewhere in the module:

- the **natural numbers** $\mathbb{N} = \{1, 2, 3, \ldots\}$
- the **integers** $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$
- the **rational numbers** $\mathbb{Q} = \{p/q \colon p \in \mathbb{Z},\ q \in \mathbb{N}\}$
- the **real numbers** $\mathbb{R}$, which can be represented by infinite decimals
- the **irrational numbers** $\mathbb{R} - \mathbb{Q}$, which are those real numbers that are not rational, such as $\sqrt{2}$, π and e
- intervals of real numbers such as $[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$, the **closed interval** with endpoints a and b, and $(a, b) = \{x \in \mathbb{R} : a < x < b\}$, the **open interval** with endpoints a and b.

In some texts the open interval is denoted by $]a, b[$.

Although in Sections 1 and 2 of this chapter we are only concerned with functions from $\mathbb{R}$ to $\mathbb{R}$, we record here the general definition of a *function* between two *arbitrary* sets. This has the advantage that we can use the same definition each time we change to a new setting.

Functions are also called *mappings* or *maps*. In some texts, the words 'mapping' and 'map' denote only continuous functions, but this is not our usage.

Definition 1.1 *Function between two arbitrary sets*

A **function** f is defined by specifying:

- a set A, the **domain** of f
- a set B, the **codomain** of f
- a **rule** $x \mapsto f(x)$ that associates with each element $x \in A$ a *unique* element $f(x) \in B$.

These are illustrated in Figure 1.1. The element $f(x)$ is called the **image** of x under f.

For $C \subseteq A$, the set $f(C) = \{f(x) : x \in C\} \subseteq B$ is called the **image set** of C under f, and for $D \subseteq B$, the set $f^{-1}(D) = \{x \in A : f(x) \in D\}$ is called the **preimage set** of D under f.

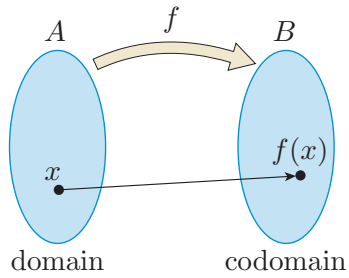


Figure 1.1 Definition of a function

There is some ambiguity in this notation in that f^{-1} does not necessarily denote the inverse function, which may not be defined, but it is used almost universally and one soon gets used to it.

Remarks

1. The preimage is also sometimes referred to as the inverse image.
2. When specifying functions symbolically we use one of three forms. For example, for the function from $\mathbb{R}$ to $\mathbb{R}$ given by $f(x) = x^2$, we use one of the following:

   $$f : \mathbb{R} \to \mathbb{R}, \quad f : x \mapsto x^2,$$
   $$f : \mathbb{R} \to \mathbb{R}, \quad f(x) = x^2,$$
   $$f(x) = x^2 \quad (x \in \mathbb{R}).$$

   In the first case we may sometimes simply write $x \mapsto x^2$ rather than $f : x \mapsto x^2$.

   Although it is good practice to include both domain and codomain when describing functions, they are occasionally omitted where the context makes them clear. This is most likely for real-valued functions of real numbers. In this case, the domain should be interpreted as the set on which the function is defined. Thus the domain of the function $f : x \mapsto \frac{1}{x}$ is $\mathbb{R} - \{0\} = \{x \in \mathbb{R} : x \neq 0\}$, since this function is defined for all non-zero real numbers.
3. In general the codomain of a function need not be the same as the image set of the domain. For example, if $f : \mathbb{R} \to \mathbb{R}$ is given by $f(x) = x^2$, then the codomain of f is $\mathbb{R}$ but the image set of $\mathbb{R}$ under f is $f(\mathbb{R}) = [0, \infty)$, the non-negative real numbers.

Exercise 1.1

Which of the following rules define functions with domain $\mathbb{Z}$ and codomain $\mathbb{N}$?

(a) Square each integer and add 1.

(b) Associate with each integer the set of all its factors.

(c) Associate with each integer the number of its factors.

Exercise 1.2

For each of those rules in Exercise 1.1 that defines a function, determine the image set of $\mathbb{N}$ and the preimage set of $\{5\}$.

1.1 The modulus function

This subsection discusses the *modulus function* – a function that will prove vital in our study of continuity for functions from $\mathbb{R}$ to $\mathbb{R}$ and which possesses some of the key properties of a distance function.

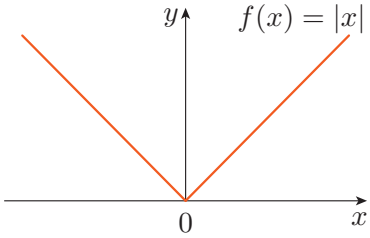


Figure 1.2 The modulus function

You may recall from Book A, Chapter 1 that the **absolute value** or **modulus** $|x|$ of a real number x is defined by

$$|x| = \begin{cases} +x, & \text{if } x \geq 0, \\ -x, & \text{if } x < 0. \end{cases}$$

For example, $|3| = 3$ and $|-3| = 3$. The graph of the modulus function is shown in Figure 1.2.

For each real number a,

$$0 \leq |a|, \quad a \leq |a|, \quad -|a| \leq a \quad \text{and} \quad |-a| = |a|. \tag{1}$$

For example, when $a = -2$,

$$0 \leq 2 = |-2|, \quad -2 \leq 2 = |-2|, \quad -|-2| = -2 \leq -2, \quad |-(-2)| = 2 = |-2|.$$

Note also that $|a| = 0$ if, and only if, $a = 0$.

Geometrically, $|a|$ gives the *distance*, without regard to direction, along the real line from a to 0.

For any real numbers a and b,

$$|ab| = |a|\,|b|, \tag{2}$$

$$|a|^2 = a^2. \tag{3}$$

For example, when $a = -2$ and $b = 3$,

$$|-2 \times 3| = |-6| = 6 = 2 \times 3 = |-2| \times |3|,$$
$$|-2|^2 = 2^2 = 4 = (-2)^2.$$

One of our main uses of the modulus function will be in *inequalities*, so we briefly discuss the solution set of the inequality

$$|x| < c,$$

where c is a constant in $\mathbb{R}$. The solution set depends on whether $c > 0$ or $c \leq 0$. For example, for $c = 2$, the set of x for which $|x| < c$ is the *open interval* $(-2, 2) = \{x\colon -2 < x < 2\} = \{x\colon |x| < 2\}$, as Figure 1.3 illustrates. However, for $c = -2$, there is no number x for which $|x| < c$, and so the solution set is the empty set, denoted $\emptyset$. In general, for $c > 0$ the solution set is the open interval $(-c, c) = \{x : -c < x < c\}$ and for $c \leq 0$ the solution set is the empty set $\emptyset$.

By *solution set* we mean the set of x for which the inequality holds.

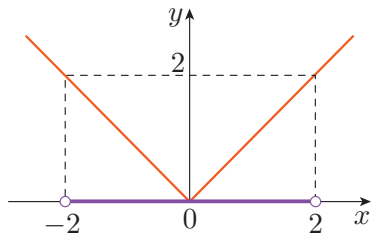


Figure 1.3 Solution set of $|x| < 2$

Thus for $c > 0$, the condition that

$$|x| < c$$

is equivalent to the double-sided inequality

$$-c < x < c.$$

This says, informally, that x lies between $-c$ and $+c$ on the real line: in other words, x is strictly less than a distance c from 0, as Figure 1.4 illustrates. The condition $-c < x < c$ is equivalent to saying that x satisfies the two simultaneous inequalities

$$-c < x \quad \text{and} \quad x < c.$$

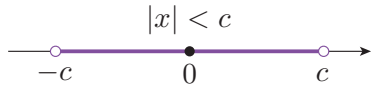


Figure 1.4 Real line showing $|x| < c$

We have similar results when the inequality is

$$|x| \leq c.$$

For $c \geq 0$, the solution set is the *closed interval* $[-c, c] = \{x\colon -c \leq x \leq c\}$. For $c < 0$, the solution set is the empty set $\emptyset$.

When $c = 0$, the solution set is $[0, 0]$, which is simply the one-point set $\{0\}$.

For $c \geq 0$, the inequality $|x| \leq c$ is equivalent both to the double-sided inequality $-c \leq x \leq c$ and to the pair of simultaneous inequalities $-c \leq x$ and $x \leq c$.

These results are summarised in the following proposition.

Proposition 1.2 *Solution sets of inequalities 1*

- For $c > 0$, the following inequalities are equivalent:
  (a) $|x| < c$, (b) $-c < x < c$, (c) $-c < x$ and $x < c$.
  Moreover, for $c \in \mathbb{R}$,
  $$\{x \in \mathbb{R}\colon |x| < c\} = \begin{cases} (-c, c), & \text{if } c > 0, \\ \emptyset, & \text{if } c \leq 0. \end{cases}$$
- For $c \geq 0$, the following inequalities are equivalent:
  (a) $|x| \leq c$, (b) $-c \leq x \leq c$, (c) $-c \leq x$ and $x \leq c$.
  Moreover, for $c \in \mathbb{R}$,
  $$\{x \in \mathbb{R} : |x| \leq c\} = \begin{cases} [-c, c], & \text{if } c \geq 0, \\ \emptyset, & \text{if } c < 0. \end{cases}$$

$[-0, 0] = \{0\}$

Exercise 1.3

Find the solution set of each of the following inequalities:

(a) $|x| < 3$; (b) $|x| \leq -3$; (c) $|3x| < 3$.

Worked Exercise 1.3

Prove that, for real numbers a and b,

$$|a| \leq |b| \quad \text{if, and only if,} \quad -|b| \leq a \leq |b|.$$

Solution

For $c \geq 0$, we have the equivalent expressions

$$|x| \leq c \quad \text{if, and only if,} \quad -c \leq x \leq c.$$

Setting $x = a$ and $c = |b|$, we obtain

$$|a| \leq |b| \quad \text{if, and only if,} \quad -|b| \leq a \leq |b|.$$

Exercise 1.4

Prove that, for real numbers a and b,

$$|a| \leq |b| \quad \text{if, and only if,} \quad -|b| \leq |a| \leq |b|.$$

We now consider more complicated inequalities.

Worked Exercise 1.4

For real numbers b and c, with $c > 0$, find the set of all real numbers x for which $|x - b| < c$.

Solution

The condition $|x - b| < c$ is equivalent to

$$-c < x - b < c,$$

which is the case if, and only if,

$$b - c < x < b + c.$$

Thus

$$\{x : |x - b| < c\} = (b - c, b + c).$$

So the required set is the open interval centred on b consisting of those points within a distance c of b, as shown in Figure 1.5.

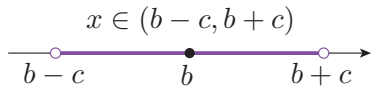


Figure 1.5 Real line showing $|x - b| < c$

A similar result holds for the inequality $|x - b| \leq c$, and so we have the following general results.

Proposition 1.5 *Solution sets of inequalities 2*

- For $b, c \in \mathbb{R}$ with $c > 0$,
$$\{x \in \mathbb{R} : |x - b| < c\} = (b - c, b + c).$$
- For $b, c \in \mathbb{R}$ with $c \geq 0$,
$$\{x \in \mathbb{R} : |x - b| \leq c\} = [b - c, b + c].$$

Exercise 1.5

Find the set of all numbers x for which:

(a) $|x - 1| < 3$; (b) $|x - 1| < 3$ and $|x + 1| < 3$.

In general, for any real numbers a and b,

$$|b - a| = |a - b|, \tag{4}$$

and this quantity gives the *distance* between a and b along the real line, as Figure 1.6 illustrates.

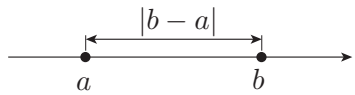


Figure 1.6 Real line showing $|b - a|$

Notice that $|b - a| = 0$ if, and only if, $a = b$.

Another important property of the modulus function is the *Triangle Inequality*.

Proposition 1.6 *Triangle Inequality in* $\mathbb{R}$

For all real numbers a and b,

$$|a + b| \leq |a| + |b|.$$

For example, with $a = -4$ and $b = 3$, $|-4 + 3| = |-1| = 1$ and $|-4| + |3| = 4 + 3 = 7$, so $|-4 + 3| \leq |-4| + |3|$.

Proof For all real numbers a and b, we have

$$\begin{aligned}|a+b|^2 &= (a+b)^2 \qquad \text{(by equation (3))}\\ &= a^2+2ab+b^2\\ &\le a^2+2|ab|+b^2 \qquad \text{(by equation (1))}\\ &= |a|^2+2\,|a|\,|b|+|b|^2 \qquad \text{(by equations (2) and (3))}\\ &= (|a|+|b|)^2.\end{aligned}$$

See Exercise 1.6.

So $|a+b|^2 \le (|a|+|b|)^2$. We now use the result that when p and q are non-negative numbers, $p^2 \le q^2$ implies $p \le q$. Hence

$$|a+b| \le |a|+|b|. \quad \blacksquare$$

Exercise 1.6

Prove that if p and q are non-negative numbers then

$$p^2 \le q^2 \implies p \le q.$$

Corollary 1.7

For all real numbers a, b and c,

$$|b-a| \le |b-c|+|c-a|.$$

Proof On noting that $b-a = (b-c)+(c-a)$, we apply Proposition 1.6 to find

$$|b-a| = |(b-c)+(c-a)| \le |b-c|+|c-a|. \quad \blacksquare$$

Exercise 1.7

Use the Triangle Inequality to show that:

(a) for all real numbers a, b and c, $|a+b+c| \le |a|+|b|+|c|$

(b) for all real numbers a and b, $|a-b| \le |a|+|b|$.

The following result provides a useful variant of the Triangle Inequality, known as the *backwards* or *reversed* form of the Triangle Inequality – or, more briefly, as the *Reverse Triangle Inequality*.

Proposition 1.8 *Reverse Triangle Inequality*

For all real numbers a and b,

$$|b-a| \ge \big||b|-|a|\big|.$$

Exercise 1.8

Prove Proposition 1.8.

1.2 Sequences

You will have met sequences of real numbers in your previous studies, but the idea of a sequence is such an important part of what follows that we now take a little time to remind you of the main ideas by recalling the notions of null sequences, convergent sequences, their combination rules and some useful theorems for determining whether a given sequence converges or diverges. We omit nearly all the proofs of these results; you will likely have seen them in an earlier module.

We begin by recalling the definition of a sequence of real numbers.

Definition 1.9 *Sequence of real numbers*

A **(real) sequence** is an unending ordered list of real numbers

$$a_1, a_2, a_3, \ldots.$$

The real number a_n ($n \in \mathbb{N}$) is the ***n*th term** of the sequence, and the whole sequence is denoted by (a_n).

A sequence (a_n) is **monotonic increasing** if for each $n \in \mathbb{N}$, $a_n \leq a_{n+1}$. It is **strictly monotonic increasing** if this inequality is strict for each n, $a_n < a_{n+1}$.

A sequence (a_n) is **monotonic decreasing** if for each $n \in \mathbb{N}$, $a_n \geq a_{n+1}$. It is **strictly monotonic decreasing** if this inequality is strict for each n, $a_n > a_{n+1}$.

Many texts use the notation $\{a_n\}$ instead of (a_n), but this makes it difficult to distinguish between the sequence (a_n) and its *set* of values, $\{a_n : n \in \mathbb{N}\}$.

Remarks

1. Alternatives to the notation (a_n) that we will sometimes use are $(a_n)_{n=1}^{\infty}$, $(a_n)_{n\in\mathbb{N}}$ (when we wish to emphasise the limits) and (in later chapters) $\mathbf{a}$.
2. By definition, unless we state otherwise, we assume that the first term of a sequence is a_1 (and not a_0).
3. The terms of a sequence (a_n) do not need to be different. For example, $-1, 1, -1, 1, \ldots$ and $1, 1, 1, 1, \ldots$ are sequences. A sequence for which all terms are the same is known as a **constant sequence**.
4. For any subset A of $\mathbb{R}$, we say that a sequence (a_n) is **in** A if $a_n \in A$ for each n. For example, the sequences $2, 4, 6, 8, \ldots$ and $1, 1, 1, 1, \ldots$ are in $\mathbb{N}$.
5. A sequence can also be viewed as a function and a more formal definition is then given by:

   a *(real) sequence* is a function $a\colon \mathbb{N} \to \mathbb{R}$ given by $n \mapsto a_n$.

We often define sequences by giving the first few terms and a formula for the nth term – for example,

$$2, 4, 6, \ldots, 2n, \ldots.$$

This sequence may also be written as

$$a_n = 2n \text{ for } n = 1, 2, 3, \ldots$$

or as $(2n)_{n=1}^{\infty}$ or simply (but possibly confusingly) as $(2n)$.

Sometimes the nth term is defined by a recursion formula, as with the Fibonacci sequence

$$1, 1, 2, 3, 5, 8, \ldots, \text{ where } a_{n+2} = a_{n+1} + a_n \text{ for } n \geq 1.$$

In this case, the first two terms of the sequence need to be specified in order to define the whole sequence.

Null sequences

If you look at the terms of the sequence

$$1, \tfrac{1}{2}, \tfrac{1}{3}, \ldots, \tfrac{1}{n}, \ldots$$

you will notice that as n gets large, the terms get closer to 0 (Figure 1.7). This sequence is an example from an important class of sequences, the class of *null sequences* – the sequences that *converge to* 0.

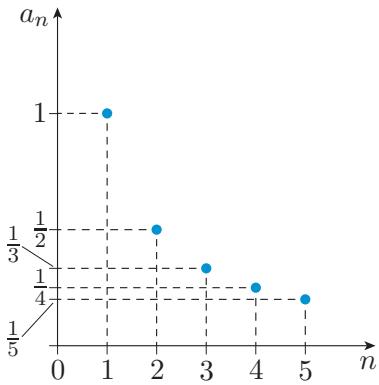


Figure 1.7 The first few terms of the sequence $\left(\frac{1}{n}\right)$

This notion of converging to 0 needs to be made precise. The idea is to show that for *any* positive number ε, we can find a natural number N such that all the terms of the sequence after a_N lie within the interval $(-\varepsilon, \varepsilon)$. A sequence is a null sequence if this property holds whatever positive number ε is taken, no matter how small.

For example, with the sequence $\left(\frac{1}{n}\right)$, suppose that $\varepsilon = 0.01$. If $\left(\frac{1}{n}\right)$ is a null sequence then there must be a value N for n for which all the subsequent terms in the sequence are within 0.01 of 0. In fact, this occurs when $N = 100$ because, from $n = 101$ onwards, $\frac{1}{n}$ is smaller than 0.01.

Often, we say simply that '(a_n) is null'.

Definition 1.10 *Null sequence*

A real sequence (a_n) is a **null sequence** if, for each $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that

$$|a_n| < \varepsilon \quad \text{whenever } n > N.$$

We say that a null sequence **converges to 0**, and write

$$a_n \to 0 \text{ as } n \to \infty, \quad \text{or simply } a_n \to 0.$$

Remarks

1. In general, for a given sequence, the value of N depends on ε; usually, the smaller ε is, the larger N will be.

2. The definition uses the absolute value $|a_n|$ because terms of the sequence may be negative. As you will recall from Subsection 1.1, $|a_n| < \varepsilon$ means that $-\varepsilon < a_n < \varepsilon$: that is, on the graph of a_n against n, the terms of a null sequence eventually lie inside a band of width 2ε centred on 0, as Figure 1.8 illustrates.

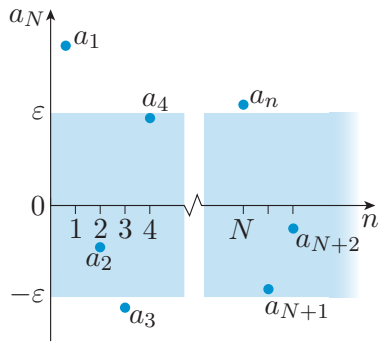


Figure 1.8 Null sequence

Worked Exercise 1.11

Prove that each of the following sequences is null:

(a) (a_n), where $a_n = 0$ for each n

(b) (b_n), where $b_n = \dfrac{1}{n^2}$ for each n

(c) (c_n), where $c_n = \dfrac{(-1)^n}{n^2}$ for each n.

Solution

(a) Let $\varepsilon > 0$ be given. Then $|a_n| = |0| = 0 < \varepsilon$ for each n.

Choose $N = 1$. Then $|a_n| < \varepsilon$ whenever $n > N$.

We deduce that (a_n) is null.

(b) Let $\varepsilon > 0$ be given. In this case $|b_n| = \left|\frac{1}{n^2}\right| = \frac{1}{n^2}$, and

$$\frac{1}{n^2} < \varepsilon \quad \text{if, and only if,} \quad \frac{1}{\sqrt{\varepsilon}} < n.$$

Choose $N = \text{int}\left(\frac{1}{\sqrt{\varepsilon}}\right)$, the largest integer less than or equal to $\frac{1}{\sqrt{\varepsilon}}$. Then, for each integer $n > N$, we have $n > \frac{1}{\sqrt{\varepsilon}}$, and hence

$$|b_n| = \frac{1}{n^2} < \varepsilon \quad \text{whenever } n > N.$$

We deduce that (b_n) is null.

(c) Let $\varepsilon > 0$ be given. Then

$$|c_n| = \left|\frac{(-1)^n}{n^2}\right| = \frac{1}{n^2} = b_n$$

for each n. Hence we choose N as in part (b), so that $|c_n| = |b_n| < \varepsilon$ whenever $n > N$.

We deduce that (c_n) is null.

Recall from Book C that $\text{int}(x)$ – also written $\lfloor x \rfloor$ – denotes the largest integer that does not exceed x. For example, $\text{int}(2) = 2$, $\text{int}(1.23) = 1$ and $\text{int}(-1.5) = -2$.

Exercise 1.9

Prove from the definition of a null sequence that the sequence (a_n) given by $a_n = \frac{1}{\sqrt{n}}$ is null.

To prove that a sequence is null, we have to show that whatever band of total width 2ε we make centred on 0, the terms of the sequence must eventually all lie inside that band. Thus to prove that a sequence (a_n) is *not* a null sequence, we need to show that this condition does not hold. This means that we need to find *just one* such band around 0 for which, no matter how far we progress in the sequence, there are always later terms of the sequence that lie outside that band.

More formally, we need to find one value of ε such that, no matter what choice of N we make, we can always find a natural number $n > N$ for which $|a_n| \geq \varepsilon$.

Worked Exercise 1.12

Prove that the sequence (a_n) given by $a_n = (-1)^n$ is not null.

Solution

Since the sequence is $-1, 1, -1, 1, \ldots$ we can take any $\varepsilon < 1$, say $\varepsilon = \frac{1}{2}$ (Figure 1.9). Then, since $|a_n| = 1$ for each n, $|a_n| \geq \varepsilon$ for each n. So there is no natural number N for which $|a_n| < \varepsilon$ for each $n > N$. Thus the sequence is not null.

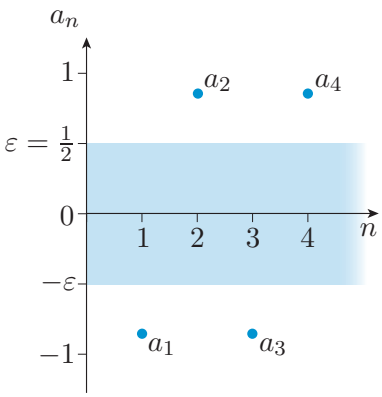


Figure 1.9 The sequence $a_n = (-1)^n$

Exercise 1.10

Prove that the constant sequence (a_n) given by $a_n = 2$ for each n is not null.

We do not prove that these sequences are null.

We record here some basic examples of null sequences.

Theorem 1.13 *Basic null sequences*

The following sequences are null:

- $\left(\frac{1}{n^p}\right)$, for any constant $p > 0$
- $(n^p c^n)$, for any constants $p \in \mathbb{R}$ and $|c| < 1$
- $\left(\frac{c^n}{n!}\right)$, for any constant $c \in \mathbb{R}$
- $\left(\frac{n^p}{n!}\right)$, for any constant $p \in \mathbb{R}$.

Examples:
$\left(\frac{1}{n^3}\right)$ when $p = 3$;
$\left(\frac{1}{2^n}\right)$ when $c = \frac{1}{2}$, $p = 0$;
$\left(\frac{2^n}{n!}\right)$ when $c = 2$;
$\left(\frac{n^{100}}{n!}\right)$ when $p = 100$.

This theorem is a consequence of the different growth rates of n^p, $\left(\frac{1}{c}\right)^n$ and $n!$ (for $|c| < 1$ and $p \in \mathbb{R}$): for n sufficiently large, it is always the case that

$$n^p \ll |c|^{-n} \ll n!.$$

It is often tricky to prove directly from the definition that a given sequence is null. However, there are several rules that allow us to combine null sequences to form new null sequences. They are summarised as follows.

Theorem 1.14 *Combination rules for null sequences*

Sum Rule	If (a_n) and (b_n) are null, then $(a_n + b_n)$ is null.
Multiple Rule	If (a_n) is null, then (λa_n) is null for each $\lambda \in \mathbb{R}$.
Product Rule	If (a_n) and (b_n) are null, then $(a_n b_n)$ is null.
Power Rule	If (a_n) is null, where $a_n \geq 0$ for each $n \in \mathbb{N}$, and $p > 0$, then (a_n^p) is null.

For example, since $\left(\frac{1}{n}\right)$ and $\left(\frac{1}{\sqrt{n}}\right)$ are (basic) null sequences, the Sum Rule implies that $\left(\frac{1}{n} + \frac{1}{\sqrt{n}}\right)$ is a null sequence, and the Product Rule (or the Power Rule) implies that $\left(\frac{1}{n^{3/2}}\right)$ is a null sequence.

Alternatively, we could note from our list of basic null sequences that $\left(\frac{1}{n^{3/2}}\right)$ is a sequence of the type $\left(\frac{1}{n^p}\right)$.

Another useful rule for proving that sequences are null is the *Squeeze Rule*. This rule uses the idea that, given a known null sequence, a sequence whose terms are eventually closer to 0 than the corresponding terms of the given sequence must also be null.

Lemma 1.15 *Squeeze Rule*

If (a_n) is null and if there is an $M \in \mathbb{N}$ such that $|b_n| \leq |a_n|$ for each $n > M$, then (b_n) is null.

Note that the Squeeze Rule can often be applied for each $n \in \mathbb{N}$ rather than for $n > M$, since if $|b_n| \leq |a_n|$ holds for each $n \in \mathbb{N}$ then it must hold for each $n > M$, whatever the value of $M \in \mathbb{N}$.

The sequence $\left(\frac{1}{2^{n^2}}\right)$ is null. This follows from the Squeeze Rule, since

$$\left|\frac{1}{2^{n^2}}\right| = \frac{1}{2^{n^2}} \leq \frac{1}{2^n} = \left|\frac{1}{2^n}\right| \quad \text{for each } n \in \mathbb{N},$$

and $\left(\frac{1}{2^n}\right)$ is a basic null sequence.

Worked Exercise 1.16

Let (a_n) be a null sequence and let $(b_n) = \left(\dfrac{a_n^2}{1+|a_n|}\right)$.
Use the combination and Squeeze Rules to prove that (b_n) is null.

Solution

We first observe that, by the Product Rule, (a_n^2) is null since it is the product of a_n with itself.

Now notice that, since $|a_n| \geq 0$ for each n,

$$\left|\frac{1}{1+|a_n|}\right| \leq 1$$

for each n, and so

$$\left|\frac{a_n^2}{1+|a_n|}\right| \leq |a_n^2|$$

for each n. Hence, by the Squeeze Rule, $(b_n) = \left(\dfrac{a_n^2}{1+|a_n|}\right)$ is null.

Exercise 1.11

Let (a_n) be a null sequence, let $\lambda \in \mathbb{R}$ with $\lambda \geq 0$ and let (b_n) be a sequence such that $|b_n| \leq \lambda |a_n|$ for each n. Use the combination and Squeeze Rules to prove that (b_n) is null.

If we were only able to investigate convergence of sequences tending to zero, then sequences would not be a very useful tool. Fortunately, we can use null sequences to define the convergence of sequences to values other than 0.

Definition 1.17 *Convergence and divergence*

A real sequence (a_n) **converges** to $l \in \mathbb{R}$ if the sequence $(a_n - l)$ is a null sequence: that is if, for each $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that

$$|a_n - l| < \varepsilon \quad \text{whenever } n > N.$$

We say that l is the **limit** of the sequence (a_n) and write $a_n \to l$ as $n \to \infty$, or simply $a_n \to l$.

A sequence that does not converge to any real number is **divergent**.

Remarks

1. A sequence converges to l if, for each $\varepsilon > 0$, the terms of the sequence eventually lie inside a band of width 2ε centred on l, as Figure 1.10 illustrates.
2. If a sequence (a_n) does not converge to l, then we write $a_n \not\to l$ as $n \to \infty$. This means that either the sequence is divergent or it converges to some number other than l.
3. Convergent real sequences have unique limits. To show this, suppose that (a_n) is a sequence that converges to the real number a and to the real number b; then both $(a_n - a)$ and $(a_n - b)$ are null sequences. Moreover, the Triangle Inequality for the modulus function tells us that for each $n \in \mathbb{N}$,

$$|a - b| = |(a - a_n) - (b - a_n)| \leq |a - a_n| + |a_n - b|.$$

But both $|a - a_n| \to 0$ and $|a_n - b| \to 0$ as $n \to \infty$ and so by the Squeeze Rule, the constant sequence $(|a - b|)$ is a null sequence, implying $|a - b| = 0$. Thus $a = b$.

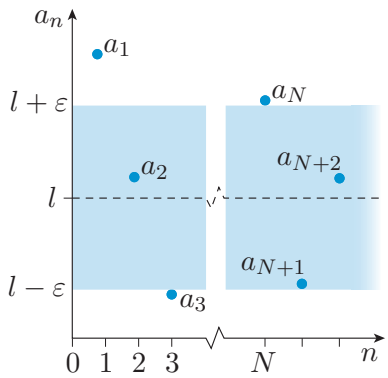


Figure 1.10 Sequence converging to l

Exercise 1.12

State conditions involving ε and N for a sequence (a_n) to be divergent, i.e. state conditions such that (a_n) does not converge to *any* limit l.

The result of Exercise 1.12 is summarised below.

(a_n) is a *divergent sequence* if for each $l \in \mathbb{R}$, there is an $\varepsilon > 0$ such that for each $N \in \mathbb{N}$, there is an $n > N$ with $|a_n - l| \geq \varepsilon$.

Worked Exercise 1.18

Prove that:

(a) the constant sequence $5, 5, 5, \ldots$ converges to 5

(b) the sequence $1, 2, 3, \ldots, n, \ldots$ is divergent.

Solution

(a) Let (a_n) be the constant sequence $5, 5, 5, \ldots$, and consider the sequence (b_n), where $b_n = a_n - 5$ for all n.

Then $b_n = 0$ for each n, and so is the constant sequence considered in Worked Exercise 1.11(a). Hence $(b_n) = (a_n - 5)$ is a null sequence, and so $a_n \to 5$ as $n \to \infty$.

(b) Here, $a_n = n$ for each n.

To show divergence, it is sufficient to find just *one* such $\varepsilon > 0$.

To prove divergence, we have to show that, given any $l \in \mathbb{R}$, there is an $\varepsilon > 0$ such that, no matter what choice of N we make, we can always find $n > N$ for which $|a_n - l| \geq \varepsilon$.

We take an arbitrary value of l and use the Reverse Triangle Inequality (Proposition 1.8) to find a lower bound on the size of $|a_n - l|$:

$$|a_n - l| = |n - l| \geq n - |l|. \tag{5}$$

For this particular sequence, we could choose any positive value we like for ε. We choose $\varepsilon = 1$ to make the algebra easier.

Let us take $\varepsilon = 1$. We will have demonstrated divergence if we can show that for any $N \in \mathbb{N}$, there is an $n > N$ for which $|a_n - l| \geq 1$, and equation (5) tells us that this will be true whenever

$$n - |l| \geq 1.$$

This inequality holds when $n \geq |l| + 1$. Thus, for each $n \geq |l| + 1$, we have

$$|a_n - l| \geq 1 = \varepsilon.$$

Here $\lceil l \rceil$ denotes the least integer greater than or equal to l.

Therefore, for each choice of N, if we take $n \in \mathbb{N}$ so that $n > \max\{N, |l| + 1\}$ (for example, take $n = \max\{N + 1, \lceil l \rceil + 1\}$, as shown in Figure 1.11), then $n > N$ and

$$|a_n - l| \geq 1 = \varepsilon.$$

We conclude that $a_n \not\to l$ as $n \to \infty$.

Since l is an arbitrary real number, (a_n) is a divergent sequence.

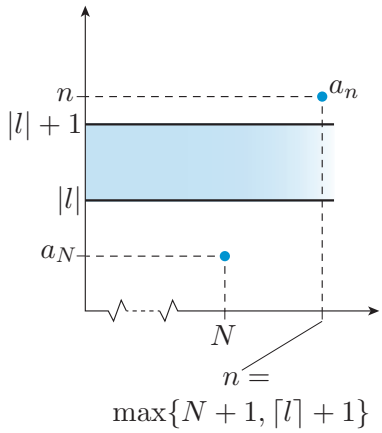


Figure 1.11 Demonstrating divergence

Remarks

1. Any constant sequence $l, l, l, \ldots$ converges to its constant value l.
2. Ideas similar to those used to show that $(n)_{n\in\mathbb{N}}$ is divergent can be used to show the following.

If (a_n) is a real null sequence, then $\left(\frac{1}{a_n}\right)$ is divergent.

Thus by taking the reciprocals of the terms of basic null sequences, we obtain divergent sequences.

As with null sequences, there are combination rules for combining convergent sequences to form new convergent sequences.

Theorem 1.19 *Combination rules for convergent sequences in* $\mathbb{R}$

If $a_n \to l$ and $b_n \to m$ as $n \to \infty$, then:

Sum Rule	$a_n + b_n \to l + m$ as $n \to \infty$
Multiple Rule	$\lambda a_n \to \lambda l$ as $n \to \infty$, for any $\lambda \in \mathbb{R}$
Product Rule	$a_n b_n \to lm$ as $n \to \infty$
Quotient Rule	$\frac{a_n}{b_n} \to \frac{l}{m}$ as $n \to \infty$, provided that $m \neq 0$
Squeeze Rule	if $l = m$ and $a_n \leq c_n \leq b_n$ for each n, then $c_n \to l$ as $n \to \infty$.

For the Quotient Rule, we are ignoring the fact that the ratio $\frac{a_n}{b_n}$ may not be defined for the first few terms; the condition that the limit of (b_n) is not zero guarantees that eventually b_n is not equal to zero.

These rules can be used, together with the basic null sequences, to prove that other sequences are convergent.

Worked Exercise 1.20

Prove that the following sequences converge, and find their limits.

(a) $(a_n) = \left((6 + 2^{-n})\left(2 - \frac{1}{n}\right)\right)$ (b) $(b_n) = \left(\frac{3n! + 2^n}{n^{50} + n!}\right)$

Solution

(a) Both (2^{-n}) and $(\frac{1}{n})$ are basic null sequences. Furthermore, by the Multiple Rule for null sequences (with $\lambda = -1$), the sequence $(-\frac{1}{n})$ is null. Also (6) and (2) are constant sequences and so converge to 6 and 2 respectively.

Hence, by the Sum Rule, $(6 + 2^{-n})$ converges to $6 + 0 = 6$ and $(2 - \frac{1}{n})$ converges to $2 + 0 = 2$.

Then, by the Product Rule, $(a_n) = ((6 + 2^{-n})(2 - \frac{1}{n}))$ converges to $6 \times 2 = 12$.

(b) Since $n!$ is the dominant term of b_n (that is, it grows faster than 2^n or n^{50} as n increases), we divide both the numerator and the denominator by $n!$. This gives

$$b_n = \frac{3 + 2^n/n!}{n^{50}/n! + 1}.$$

Both $\left(\frac{2^n}{n!}\right)$ and $\left(\frac{n^{50}}{n!}\right)$ are basic null sequences, and (3) and (1) are constant sequences. So, by the Sum Rule, the sequence $\left(3 + \frac{2^n}{n!}\right)$ converges to 3 and the sequence $\left(\frac{n^{50}}{n!} + 1\right)$ converges to 1. Hence, by the Quotient Rule, (b_n) converges to $\frac{3}{1} = 3$.

For sequences whose general term is expressed as a quotient, the technique of dividing both the numerator and the denominator by the *dominant term* is often useful when trying to prove convergence or divergence.

For constants c and p, $n!$ dominates c^n, which in turn dominates n^p.

Exercise 1.13

Prove that the following sequence converges, and find its limit.

$$\left(\frac{4n^2 + 1000^n}{n! + 7n^3}\right)$$

So far, in order to prove a given sequence is convergent, we have had to identify its limit first and then prove that the sequence converges to this limit. Sometimes this is difficult to do. For example, it is not obvious that the limit of the sequence (a_n) given by

$$a_n = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2}$$

is $\frac{\pi^2}{6}$. So it is useful to have results that guarantee that a given sequence is convergent, even if we do not know its limit.

We now record a very useful result for concluding that a sequence is convergent, the *Monotone Convergence Theorem* for sequences, that does not require us to identify the limit. It is a powerful tool and we will return to the general problem of how to identify convergent sequences without finding their limits in a later chapter.

We expect that you will likely have met this result before and we do not prove it here.

See Book F, Chapter 23, *Completeness*.

To introduce it, we first remind you of the terms *upper bound* and *lower bound*.

Definition 1.21 *Upper and lower bounds*

Let A be a subset of real numbers. Then $b_l \in \mathbb{R}$ is a **lower bound** of A if for each $a \in A$, $b_l \leq a$.

Similarly, $b_u \in \mathbb{R}$ is an **upper bound** of A if for each $a \in A$, $a \leq b_u$.

Notice that there is no assumption in this definition that $b_l, b_u \in A$, nor does this terminology assume that these are the best possible choices of bound.

Theorem 1.22 *Monotone Convergence Theorem*

If (a_n) is a monotone increasing sequence that is bounded above by a real number b_u, then (a_n) is convergent with limit $l \leq b_u$.

If (a_n) is a monotone decreasing sequence that is bounded below by a real number b_l, then (a_n) is convergent with limit $l \geq b_l$.

Figure 1.12 illustrates these upper and lower bounds.

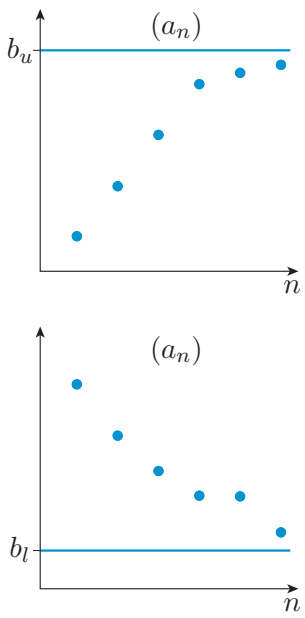


Figure 1.12 Upper and lower bounds for (a_n)

Worked Exercise 1.23

Prove that the sequence (a_n) given by

$$a_n = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2}$$

is convergent.

Solution

It is certainly the case that (a_n) is a monotonic increasing sequence, since for each $n \in \mathbb{N}$,

$$a_{n+1} - a_n = \frac{1}{(n+1)^2} > 0.$$

So if we can show that (a_n) is bounded above, then we can use the Monotone Convergence Theorem to conclude that (a_n) is a convergent sequence. But for each $n \geq 2$,

$$\frac{1}{n^2} \leq \frac{1}{n(n-1)} = \frac{1}{n-1} - \frac{1}{n}$$

and so for $n \geq 2$,

$$\begin{aligned} a_n &= 1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} \\ &\leq 1 + \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \cdots + \left(\frac{1}{n-1} - \frac{1}{n}\right) \\ &= 2 - \frac{1}{n}. \end{aligned}$$

$a_1 = 1 \leq 2.$

Hence (a_n) is a sequence that is bounded above by 2. We conclude by the Monotone Convergence Theorem that (a_n) is a convergent sequence with limit at most 2.

Next we recall a technique for investigating whether a sequence is divergent: looking at appropriate *subsequences*. The essential features of a subsequence are that all of its terms are in the original sequence, and that the terms occur in the same order as in the original sequence. In general, given a sequence (a_n), we can form many different subsequences.

Definition 1.24 *Subsequence*

The sequence $(a_{n_k})_{k=1}^{\infty}$ is a **subsequence** of the sequence (a_n) if $(n_k)_{k=1}^{\infty}$ is a strictly increasing sequence of positive integers, i.e. if $1 \leq n_1 < n_2 < n_3 < \cdots$.

In particular:

- when $n_k = 2k$, we have $(a_{2k})_{k=1}^{\infty}$, which is the **even subsequence** of (a_n)
- when $n_k = 2k - 1$, we have $(a_{2k-1})_{k=1}^{\infty}$, which is the **odd subsequence** of (a_n).

Remarks

1. Every sequence is a subsequence of itself, obtained by putting $n_k = k$, for $k = 1, 2, 3, \ldots$.
2. The even subsequence (a_{2k}) consists of the even-numbered terms $a_2, a_4, a_6, \ldots$ of the original sequence. For example, for the sequence $(a_n) = 1, 4, 9, 16, 25, 36, \ldots, n^2, \ldots$, the even subsequence (a_{2k}) is $4, 16, 36, \ldots, (2k)^2, \ldots$. Similarly, the odd subsequence consists of the odd-numbered terms $a_1, a_3, a_5, \ldots$.

Exercise 1.14

Show that (2^{k^2}) is a subsequence of (2^n).

The terms of (2^{k^2}) are $2, 2^4, 2^9, 2^{16}, \ldots$.

The following result relates the convergence of a given sequence to the convergence of its subsequences and provides us with a way to show that a sequence is divergent.

Proposition 1.25 *Convergence of subsequences*

If (a_n) is a convergent sequence with limit l, then every subsequence (a_{n_k}) is convergent with the same limit l.

Proposition 1.25 can be used to show that a given sequence is *divergent*, since it tells us the following.

1. If (a_n) has a divergent subsequence, then (a_n) is itself divergent.
2. If a sequence (a_n) has two convergent subsequences whose limits are different, then (a_n) is divergent.

Worked Exercise 1.26

Show that $(a_n) = ((-1)^n)$ is a divergent sequence.

Solution

We produce two subsequences of (a_n) that converge to different limits.

Consider the even subsequence $(a_{2k}) = ((-1)^{2k})$. Each term of this subsequence is 1, and so this subsequence converges to 1.

Now consider the odd subsequence $(a_{2k-1}) = ((-1)^{2k-1})$. Each term is -1, and so this subsequence converges to -1, a different limit. (See Figure 1.13.)

By Proposition 1.25, we conclude that $(a_n) = ((-1)^n)$ is divergent.

In Worked Exercise 1.12 we proved that this sequence is not a null sequence.

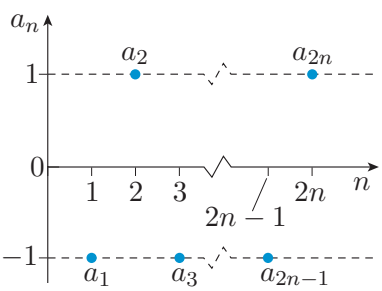


Figure 1.13 Subsequences of $(a_n) = ((-1)^n)$

Exercise 1.15

Determine whether the following sequence is convergent or divergent, justifying your answer.

$$\left((-1)^n \frac{1-n}{n}\right)$$

The final exercise in this subsection is an opportunity for you to practise testing sequences for convergence.

Exercise 1.16

Determine whether each of the following sequences converges. If so, what is its limit?

(a) $\left(\frac{3n^3}{n!} + \frac{7}{n^2}\right)$; (b) $\left((n+1)^2 - n^2\right)$; (c) $\left(\sqrt{n+1} - \sqrt{n}\right)$.

Hint: for part (c), you may find the following identity helpful:

$$a - b = (\sqrt{a} - \sqrt{b})(\sqrt{a} + \sqrt{b}) \text{ for } a, b \geq 0.$$

2 Continuity on the real line

In this section, we recall the definition of a continuous function from $\mathbb{R}$ to $\mathbb{R}$. We use (convergent) sequences as our main tool for testing the continuity of functions.

2.1 Continuous and discontinuous functions

Consider the functions $f_1 : \mathbb{R} \to \mathbb{R}$ and $f_2 : \mathbb{R} \to \mathbb{R}$ illustrated in Figure 2.1 and given by

$$f_1(x) = x^2,$$

$$f_2(x) = \begin{cases} 1, & \text{if } x > 0, \\ 0, & \text{if } x \leq 0. \end{cases}$$

Intuitively, we would say that f_1 is continuous since its graph has no jumps or breaks in it, whereas f_2 is discontinuous since there is a jump at $x = 0$. We could go on to say that f_1 is continuous at every point $x \in \mathbb{R}$, whereas f_2 is discontinuous at $x = 0$.

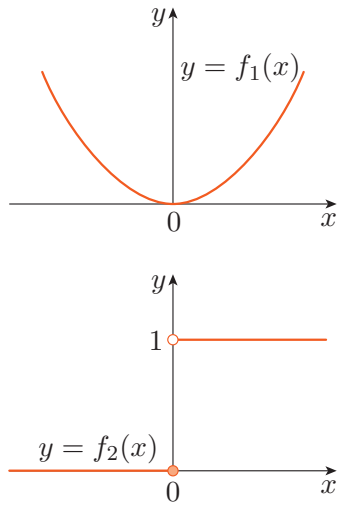


Figure 2.1 The functions $f_1(x)$ and $f_2(x)$

Our immediate objective is to capture this intuitive idea of a continuous function in a precise way. We do this by using the idea that if f is continuous at a point a and if x is a point near to a, then $f(x)$ should be near to $f(a)$. In other words, our definition should say in a precise way that

if x tends to a, then $f(x)$ tends to $f(a)$.

We make this notion exact by using the convergence of sequences.

Definition 2.1 *Continuity at a point*

Let $A \subseteq \mathbb{R}$ and let $f\colon A \to \mathbb{R}$ be a function.

Then f is **continuous** at $a \in A$ if:

whenever (x_n) is a sequence in A for which $x_n \to a$ as $n \to \infty$, then the sequence $(f(x_n))$ converges to $f(a)$.

If f does not satisfy this condition at $a \in A$ – that is, there is a sequence (x_n) in A for which $x_n \to a$ as $n \to \infty$ but $f(x_n)$ does not converge to $f(a)$ – then we say that f is **discontinuous** at a.

Remarks

1. This definition of continuity is often called the **sequential definition of continuity**, or simply **sequential continuity**, to distinguish it from alternative definitions.
2. To show that f is continuous at a, we must show that $(f(x_n))$ converges to $f(a)$ for *every possible choice* of a sequence (x_n) that converges to a.
3. If the sequence $(f(x_n))$ converges to $f(a)$ for a *particular* sequence (x_n) that converges to a (in the domain A), then we often write

   $$f(x_n) \to f(a) \text{ as } x_n \to a.$$

   If the sequence $(f(x_n))$ converges to $f(a)$ for *every* sequence (x_n) that converges to a (in the domain A), then we often write

   $$f(x) \to f(a) \text{ as } x \to a.$$

Continuity as defined above is the property of a function *at a point*. It is easily extended to define continuity *on a set*.

Definition 2.2 *Continuity on a set*

Let $S \subseteq A \subseteq \mathbb{R}$ and let $f\colon A \to \mathbb{R}$ be a function.

We say that f is **continuous** on S when f is continuous at each point in S.

We say that f is **discontinuous** on S if f is discontinuous at at least one point in S.

Remarks

1. If we simply say that a function $f: A \to \mathbb{R}$ is *continuous* (or *discontinuous*), by implication we mean that it is continuous (or discontinuous) on A.
2. To prove continuity on a set S, we need to show that the function f is continuous at *each* point a in S – if f is a 'hybrid' function made up of two or more functions, then this may require different arguments depending on the value of a.
3. To prove discontinuity on a set S, all that is needed is to prove it for a *single*, *specific* point in S.

We now show how the definition of continuity is used to prove that a function is continuous.

Worked Exercise 2.3

Prove that the function $f_1: \mathbb{R} \to \mathbb{R}$ given by

$$f_1(x) = x^2$$

is (a) continuous at 2 and (b) continuous on $\mathbb{R}$.

Solution

(a) The definition of continuity tells us that we must show that whenever (x_n) is a sequence in $\mathbb{R}$ that converges to 2, then the sequence $(f_1(x_n))$ converges to $f_1(2) = 4$.

Let (x_n) be a sequence (in $\mathbb{R}$) that converges to 2. We know that $f_1(x_n) = x_n^2$. Therefore, by the Product Rule for sequences,

$$f_1(x_n) = x_n \times x_n \to 2 \times 2 = 4 \text{ as } n \to \infty.$$

But $f_1(2) = 4$, and so $f_1(x_n) \to f_1(2)$ as $x_n \to 2$.

Hence f_1 is continuous at 2.

(b) This is a generalisation of the method in part (a).

Let a be a general point in $\mathbb{R}$. By the definition of continuity we must show that whenever (x_n) is a sequence in $\mathbb{R}$ that converges to a, then the sequence $(f_1(x_n))$ converges to $f_1(a) = a^2$.

Let (x_n) be a sequence (in $\mathbb{R}$) that converges to a. By the Product Rule for sequences,

$$f_1(x_n) = x_n^2 = x_n \times x_n \to a \times a = a^2 \text{ as } n \to \infty.$$

But $f_1(a) = a^2$, and so $f_1(x_n) \to f_1(a)$ as $x_n \to a$.

Hence f_1 is continuous at a.

Since a is an arbitrary point of $\mathbb{R}$, f_1 is continuous on $\mathbb{R}$.

Exercise 2.1

Prove that the function $f\colon \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = x^3 + 3$$

is continuous on $\mathbb{R}$.

To show that a function f is *discontinuous* at a point $a \in A$, we just need to find a single sequence (x_n) in A for which

$$x_n \to a \text{ but } f(x_n) \not\to f(a) \text{ as } n \to \infty.$$

Worked Exercise 2.4

Prove that the function $f_2\colon \mathbb{R} \to \mathbb{R}$ given by

$$f_2(x) = \begin{cases} 1, & \text{if } x > 0, \\ 0, & \text{if } x \leq 0, \end{cases}$$

is discontinuous at 0.

Solution

The definition of discontinuity tells us that we must find a sequence (x_n) in $\mathbb{R}$ for which $x_n \to 0$ as $n \to \infty$, but $f_2(x_n) \not\to f_2(0)$ as $n \to \infty$.

Since $f_2(0) = 0$ and $f_2(x) = 1$ for $x > 0$, we choose a sequence (x_n) that converges to 0 'from the right', so that $f_2(x_n) = 1$ for each n. One such sequence is given by $(x_n) = (\frac{1}{n})$, as shown in Figure 2.2. For this choice of sequence, we have

$$x_n = \tfrac{1}{n} \to 0 \text{ as } n \to \infty \text{ and yet } f_2(x_n) = 1 \text{ for each } n,$$

so $f_2(x_n) \to 1$ as $n \to \infty$. Since $f_2(0) = 0$ and $f_2(x_n) \not\to 0$,

$$f_2(x_n) \not\to f_2(0) \text{ as } n \to \infty,$$

and we conclude that f_2 is discontinuous at 0.

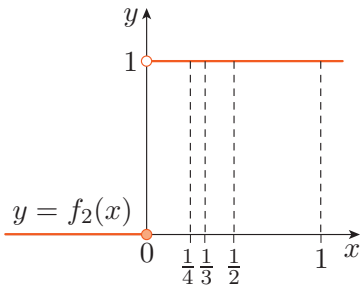


Figure 2.2 The sequence $(x_n) = (\frac{1}{n})$

Thus f_2 is discontinuous only at $a = 0$. This is what we would expect from looking at its graph, given in Figure 2.2.

Exercise 2.2

Prove that f_2 is continuous at a for each $a \neq 0$.

Worked Exercise 2.5

Let $f\colon \mathbb{R} \to \mathbb{R}$ be given by the rule

$$f(x) = \begin{cases} 1, & \text{if } x \in \mathbb{Q}, \\ 0, & \text{if } x \notin \mathbb{Q}. \end{cases}$$

Show that f is discontinuous at every $a \in \mathbb{R}$.

Solution

Let $a \in \mathbb{R}$.

We show that f is discontinuous at a by finding a sequence (x_n) such that

$$x_n \to a, \text{ but } f(x_n) \not\to f(a) \text{ as } n \to \infty.$$

If a is *rational*, then by definition $f(a) = 1$. We take a sequence of irrational numbers that converges to a – for example, let $x_n = a + \frac{1}{n}\sqrt{2}$. Then the combination rules for convergent sequences tell us that $x_n \to a + 0 \times \sqrt{2} = a$ as $n \to \infty$ but $f(x_n) = 0 \not\to f(a) = 1$, as $n \to \infty$.

We know that $a + \frac{1}{n}\sqrt{2}$ is irrational since if $a + \frac{1}{n} = \frac{p}{q}$ for some choice of integers p and q (with $q \neq 0$), then rearranging this expression would give $\sqrt{2} = n(\frac{p}{q} - a)$, a rational number.

Hence f is not continuous at a if $a \in \mathbb{Q}$.

If a is *irrational*, then $f(a) = 0$, and we choose a sequence of *rational* numbers (x_n) that converges to a. (Such a sequence exists because the density property of the reals tells us that between any two distinct real numbers there is a rational number; in particular, we can find a rational number x_n in the interval $(a, a + \frac{1}{n})$.)

The Squeeze Rule then guarantees that the sequence of rationals (x_n) converges to a, since $a + \frac{1}{n} \to a$ and $a \to a$ as $n \to \infty$.

Thus $x_n \to a, \text{ but } f(x_n) = 1 \not\to f(a) = 0, \text{ as } n \to \infty.$

Hence f is not continuous at a if $a \notin \mathbb{Q}$.

We deduce that f is discontinuous at every point in $\mathbb{R}$.

Let us now see what our definition tells us about the continuity of a function whose behaviour is less obvious.

Worked Exercise 2.6

Let $f\colon \mathbb{R} \to \mathbb{R}$ be the function defined by

$$f(x) = \begin{cases} \cos(\pi/x), & \text{if } x \neq 0, \\ 0, & \text{if } x = 0. \end{cases}$$

Decide whether f is continuous at 0, and justify your decision.

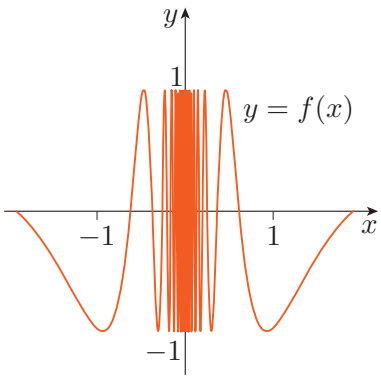


Figure 2.3 The graph of f for Worked Exercise 2.6

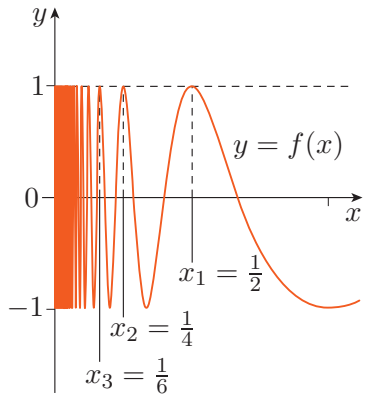


Figure 2.4 Choosing the sequence (x_n)

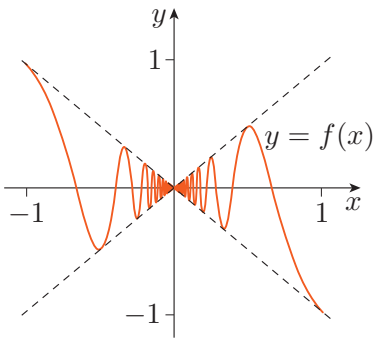


Figure 2.5 The graph of f for Exercise 2.3

Solution

As Figure 2.3 shows, the graph of f oscillates very rapidly close to the origin, and there are points x, arbitrarily close to the origin, where $\cos(\pi/x)$ seems to take values that are far from $0 = f(0)$. We expect that f is *discontinuous* at 0.

In order to prove this, we need to find a single sequence (x_n) in $\mathbb{R}$ for which

$$x_n \to 0 \text{ but } f(x_n) \not\to f(0) = 0.$$

We choose a sequence (x_n) of points for which $\cos(\pi/x_n) = 1$ (that is, the maximum points of the oscillation), so that $f(x_n) = \cos(\pi/x_n) = 1$ for each n and hence $f(x_n) \to 1 \neq 0$.

Now $\cos a = 1$ whenever a is of the form $2\pi n$, where n is an integer. It follows that $\cos(\pi/x) = 1$ whenever $\pi/x = 2\pi n$, i.e. whenever $x = \frac{1}{2n}$. So if we take (x_n) to be the sequence given by $x_n = \frac{1}{2n}$, as shown in Figure 2.4, then

$$f(x_n) = \cos\left(\frac{\pi}{x_n}\right) = \cos\left(\frac{\pi}{1/(2n)}\right) = \cos 2\pi n = 1 \quad \text{for each } n.$$

Moreover, $x_n \to 0$ as $n \to \infty$.

So we have found a sequence (x_n) for which $x_n \to 0$ but $f(x_n) \not\to f(0)$.

Thus f is discontinuous at 0.

Exercise 2.3

Prove that the function $f\colon \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} x\cos(\pi/x), & \text{if } x \neq 0, \\ 0, & \text{if } x = 0, \end{cases}$$

is continuous at 0. (This function is shown in Figure 2.5.)

Exercise 2.4

Let $f\colon \mathbb{R} - \{1\} \to \mathbb{R}$ be given by $f(x) = \dfrac{x^2 - 1}{x - 1}$.

(a) Suppose that (x_n) is a sequence in $\mathbb{R} - \{1\}$ with $x_n \to 1$ as $n \to \infty$. Show that $(f(x_n))$ is a convergent sequence and find its limit.

(b) Deduce that if $g\colon \mathbb{R} \to \mathbb{R}$ is given by

$$g(x) = \begin{cases} f(x), & \text{if } x \neq 1, \\ 2, & \text{if } x = 1, \end{cases}$$

then g is continuous at $a = 1$.

The discontinuity at $a = 1$ for the function f in Exercise 2.4 is called a *removable* discontinuity because, by making an appropriate change to the definition of f (as in Exercise 2.4(b)), it is possible to define a continuous function that is identical to f everywhere except at the point $a = 1$. This contrasts with the situation for the function f of Worked Exercise 2.6, where the point $a = 0$ is known as an *essential* discontinuity – it is not possible to change the definition of this function at $a = 0$ in such a way as to make it continuous there. (We do not prove this, but the argument is similar to that given in Worked Exercise 2.6.)

2.2 Rules for continuous functions

It would be very laborious if every time we wished to show a given function was continuous, we had to go back to 'first principles' and work directly from the definition of continuity. Fortunately, by using a list of standard continuous functions together with combination rules for continuous functions (that are similar to those we have for convergent sequences), we can frequently avoid recourse to direct use of the definition of continuity.

We start by recording the combination rules for continuous functions from $\mathbb{R}$ to $\mathbb{R}$. They follow directly from the combination rules for convergent sequences in $\mathbb{R}$.

This should not be a surprise: continuity is defined in terms of convergent sequences, so one would expect many results about convergent sequences to have analogues for continuous functions.

Combination rules for continuous functions from $\mathbb{R}$ to $\mathbb{R}$

Let $A \subseteq \mathbb{R}$ and suppose $f\colon A \to \mathbb{R}$ and $g\colon A \to \mathbb{R}$ are continuous at $a \in A$.

Then the following functions are also continuous at a:

Sum Rule	$f + g\colon A \to \mathbb{R}$, defined by $(f + g)(x) = f(x) + g(x)$
Multiple Rule	$\lambda f\colon A \to \mathbb{R}$ where $\lambda \in \mathbb{R}$, defined by $(\lambda f)(x) = \lambda \times f(x)$
Product Rule	$fg\colon A \to \mathbb{R}$, defined by $(fg)(x) = f(x)g(x)$
Quotient Rule	$f/g\colon A - \{x : g(x) = 0\} \to \mathbb{R}$, defined by $(f/g)(x) = f(x)/g(x)$, provided $g(a) \neq 0$.

Another useful result tells us that restricting the domain of a continuous function to a smaller set still gives us a continuous function. Before we state this result, we need a definition.

Definition 2.7 *Restricted function*

If the domain of a function $f\colon A \to B$ is restricted to a set C, where $C \subseteq A$, then the resulting **restricted function**, known as the **restriction of f to C**, is denoted by $f|_C\colon C \to B$, and given by $f|_C(x) = f(x)$ for each $x \in C$.

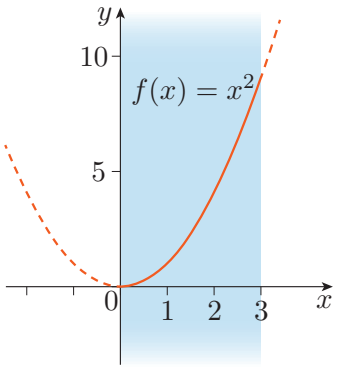


Figure 2.6 The function $f(x) = x^2$ restricted to $[0, 3]$

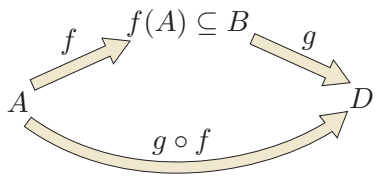


Figure 2.7 Composition

Here $A = B = C = \mathbb{R}$, $D = [-1, 1]$ and $f(A) = \mathbb{R} \subseteq B$.

Proposition 2.8 *Restriction Rule for continuous functions on $\mathbb{R}$*

Let $A \subseteq \mathbb{R}$, let $f\colon A \to \mathbb{R}$ and let $B \subseteq A$. If f is continuous at $b \in B$, then the restricted function $f|_B\colon B \to \mathbb{R}$ is continuous at b.

In particular, if f is continuous on A, then $f|_B$ is continuous on B.

For example, the function $f(x) = x^2$ is continuous on $\mathbb{R}$. Let $B = [0, 3]$. Then $f|_B\colon B \to \mathbb{R}$ is f restricted to the interval $[0, 3]$ and is continuous on $[0, 3]$, as Figure 2.6 illustrates.

Our final useful result about combining continuous functions allows us to compose continuous functions to form another continuous function. Before stating this result, we remind you of the definition of the *composition* of two functions.

Definition 2.9 *Composition*

If $f\colon A \to C$ and $g\colon B \to D$ are functions with $f(A) \subseteq B$, then the **composed function** or **composition** $g \circ f\colon A \to D$ is given by $(g \circ f)(x) = g(f(x))$.

Note that the condition $f(A) \subseteq B$ is needed so that the image set of the function f lies inside the domain of g (Figure 2.7). Some definitions of composition omit this condition, but we will require it to hold.

For example, let $f\colon \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^3$ and let $g\colon \mathbb{R} \to [-1, 1]$ be given by $g(x) = \sin x$. Then $g \circ f\colon \mathbb{R} \to [-1, 1]$ is given by $(g \circ f)(x) = \sin(x^3)$.

Proposition 2.10 *Composition Rule*

Let A, $B \subseteq \mathbb{R}$ and let $f\colon A \to \mathbb{R}$ and $g\colon B \to \mathbb{R}$ be continuous on A and B respectively, with $f(A) \subseteq B$. Then the composed function $g \circ f\colon A \to \mathbb{R}$ is continuous on A.

We prove a more general version of this result in Chapter 14 (Proposition 4.4) and do not prove it here.

These rules enable us to use a collection of basic continuous functions to form new functions that must then also be continuous on their domains.

Basic continuous functions

The following functions are continuous:

- polynomials and rational functions
- $f(x) = |x|$
- $f(x) = \sqrt{x}$
- the trigonometric functions (sine, cosine and tangent) and their inverses (restricted to appropriate domains)
- the logarithmic function
- the exponential function.

Note that by convention, the domains of these basic continuous functions are taken to be the largest subset of the reals for which the functions are defined. For example, the domain of a rational function excludes only those points where the denominator is zero, the domain of $f(x) = \sqrt{x}$ is $[0, \infty)$, and the domain of tan excludes the points $(2n + 1)\pi/2$ for all $n \in \mathbb{Z}$.

The application of the combination rules to functions constructed from basic continuous functions is illustrated in the following worked exercise.

Worked Exercise 2.11

Using the above rules for continuous functions, show that the function $f\colon [-\pi, \pi] \to \mathbb{R}$ given by

$$f(x) = e^x + 3\cos 2x$$

is continuous on $[-\pi, \pi]$.

Solution

The functions $x \mapsto 2x$, $x \mapsto e^x$ and $x \mapsto \cos x$ are all basic continuous functions, defined on the whole of $\mathbb{R}$.

By the Restriction Rule, both $x \mapsto e^x$ and $x \mapsto 2x$ are continuous functions on the interval $[-\pi, \pi]$.

The image set of $x \mapsto 2x$ restricted to $[-\pi, \pi]$ is $[-2\pi, 2\pi] \subseteq \mathbb{R}$. Thus, by the Composition Rule, $x \mapsto \cos 2x$ is continuous on $[-\pi, \pi]$. So, by the Multiple Rule, $x \mapsto 3\cos 2x$ is continuous on $[-\pi, \pi]$.

Finally, by the Sum Rule, f is continuous on $[-\pi, \pi]$.

Exercise 2.5

Show that $f\colon (0, \infty) \to \mathbb{R}$ given by

$$f(x) = 2\sin(1/x) - 3e^{-4x^3}$$

is continuous on $(0, \infty)$.

Exercise 2.6

Let $f: \mathbb{R} \to \mathbb{R}$ be a continuous function. Show that the function $|f|: \mathbb{R} \to \mathbb{R}$ given by $|f|(x) = |f(x)|$ is continuous on $\mathbb{R}$.

2.3 Continuity theorems

In this subsection we remind you of three important theorems concerning continuous functions from $\mathbb{R}$ to $\mathbb{R}$. We do not prove these theorems here because we expect that you will have encountered them in your earlier studies. These theorems give us important information about the properties of continuous functions from $\mathbb{R}$ to $\mathbb{R}$. One of our goals when we extend the notion of continuity to more general settings will be to see how far these theorems still hold true.

The proofs of the more general results we obtain in later chapters will also give proofs of the results we discuss here.

All three theorems concern continuous functions whose domain is a **closed bounded interval**, that is, an interval of the form

$$[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$$

for real numbers a and b.

The Intermediate Value Theorem

Our first theorem says that continuous functions do not 'omit any values', and so there cannot be any jumps or gaps in their graphs. Thus, when f is continuous on a closed interval and attains two particular values, it must also take all the values between them. For example, let f be continuous on $[a, b]$, and suppose that $f(a) < 0$ and $f(b) > 0$. Then the theorem asserts that there must be at least one point $c \in (a, b)$ for which $f(c) = 0$ – that is, the graph of f must cross the x-axis, as Figure 2.8 illustrates.

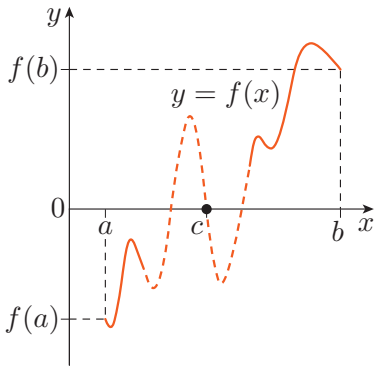


Figure 2.8 Graph of continuous function on $[a, b]$

Theorem 2.12 *Intermediate Value Theorem*

Let $f: [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ with $f(a) \neq f(b)$ and let k be any number lying between $f(a)$ and $f(b)$. Then there exists a number $c \in (a, b)$ such that $f(c) = k$.

Remarks

1. Notice that the theorem says nothing about the case when $f(a) = f(b)$.
2. By making use of the Restriction Rule (Proposition 2.8), this theorem can be adapted to apply to all intervals and not just to ones that are closed and bounded. To see this, suppose g is a continuous function from an interval I to $\mathbb{R}$, and a and b lie in I (with $a < b$). Then, by the Restriction Rule, $f: [a, b] \to \mathbb{R}$ given by $f(x) = g(x)$ for $x \in [a, b]$ is a continuous function on $[a, b]$ and so satisfies the hypotheses of the

I is of one of the following forms: $\mathbb{R}$, (r, s), $[r, s)$, $(r, s]$, $[r, s]$, $(-\infty, s)$, $(-\infty, s]$, (r, ∞), $[r, \infty)$ for $r, s \in \mathbb{R}$.

Intermediate Value Theorem. We conclude that if $g\colon I \to \mathbb{R}$ is continuous on the interval I and $a, b \in I$ are given, then whenever k lies between $g(a)$ and $g(b)$, there is $c \in (a, b)$ such that $g(c) = k$.

3. The Intermediate Value Theorem does not hold if the domain of f is not an interval. For example, consider the function $f(x) = \frac{1}{x}$ restricted to the set $[-1, 1] - \{0\}$, as shown in Figure 2.9. This function is continuous on $[-1, 1] - \{0\}$ and has $f(-1) < 0$ and $f(1) > 0$, but there is no value of $x \in [-1, 1] - \{0\}$ for which $f(x) = 0$.

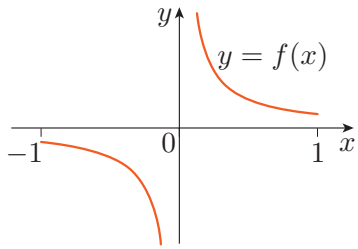


Figure 2.9 Graph of $f(x) = \frac{1}{x}$ on $[-1, 1] - \{0\}$

Worked Exercise 2.13

Let $f\colon [-\pi, \pi] \to \mathbb{R}$ be the function given by

$$f(x) = e^x + 3\cos 2x.$$

Prove that there is at least one point $c \in [-\pi, \pi]$ for which $f(c) = 0$.

Solution

We saw in Worked Exercise 2.11 that f is continuous on $[-\pi, \pi]$.

To show that there is a point $c \in [-\pi, \pi]$ for which $f(c) = 0$, using the Intermediate Value Theorem, we need to find an interval $[a, b]$ for which $f(a) < 0$ and $f(b) > 0$. We cannot use the endpoints of $[-\pi, \pi]$ because both $f(-\pi)$ and $f(\pi)$ are positive (see Figure 2.10). So we need to restrict our attention to a suitable subinterval of $[-\pi, \pi]$. The set $[-\frac{\pi}{2}, 0]$ is a closed interval contained in $[-\pi, \pi]$, and the Restriction Rule implies that f is continuous on it. Moreover,

$$f(-\tfrac{\pi}{2}) = e^{-\frac{\pi}{2}} + 3\cos(-\pi) = e^{-\frac{\pi}{2}} - 3 < 0,$$

$$f(0) = e^0 + 3\cos 0 = 1 + 3 = 4 > 0.$$

Since the conditions of the Intermediate Value Theorem are satisfied by f on $[-\frac{\pi}{2}, 0]$, we deduce that there is a number c (with $-\frac{\pi}{2} < c < 0$) for which $f(c) = 0$.

We could also have used the interval $[-\pi, -\frac{\pi}{2}]$, as can be seen from Figure 2.10.

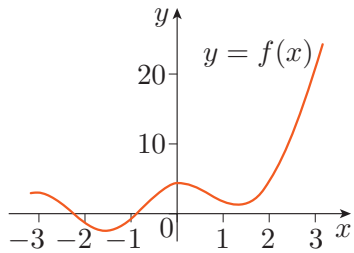


Figure 2.10 Graph of $f(x) = e^x + 3\cos 2x$ on $[-\pi, \pi]$

$e^{-\frac{\pi}{2}} < e^0 = 1.$

The Intermediate Value Theorem is an *existence result*: it tells us that there is a point in the interval whose image under f is 0, but it does not tell us how to find such a point. Thus, although we can deduce that the function $f\colon [-\pi, \pi] \to \mathbb{R}$ given by $f(x) = e^x + 3\cos 2x$ takes the value 0 at some point(s) in the interval $[-\frac{\pi}{2}, 0]$, we must use different methods to find the value(s) of x for which this is true.

The unique solution to $e^x + 3\cos 2x = 0$ in $[-\frac{\pi}{2}, 0]$ is approximately $x = -0.8564$.

Although the theorem does not provide a method for finding a point c at which a function f takes an intermediate value k, it is nevertheless very useful. It enables us to know in advance whether such a point c exists, and it applies to *all* continuous functions defined on a closed bounded interval.

Exercise 2.7 *A fixed-point theorem*

Let $f\colon [0,1] \to [0,1]$ be a continuous function. Use the Intermediate Value Theorem to show that there is a number $k \in [0,1]$ such that $f(k) = k$.

Hint: consider the function $g\colon [0,1] \to \mathbb{R}$ defined by $g(x) = f(x) - x$.

Exercise 2.7 shows that each continuous function $f\colon [0,1] \to [0,1]$ has a *fixed point*, a point that maps to itself under f.

The Boundedness and Extreme Value Theorems

The next two theorems involve not only closed bounded intervals but *boundedness* on $\mathbb{R}$ more generally. So, before we discuss the theorems, we remind you of the definition of boundedness on $\mathbb{R}$.

Definition 2.14 *Boundedness*

A set $A \subset \mathbb{R}$ is **bounded** if there exists a real number M such that $|x| \leq M$ for each $x \in A$.

A function $f\colon A \to \mathbb{R}$ is **bounded** on A if there is a real number M such that $|f(x)| \leq M$ for each $x \in A$.

Notice that closed bounded intervals $[a,b]$ are bounded sets.

Open intervals of the form (a,b) for real numbers a and b are also bounded sets.

Our first theorem is an immediate consequence of the second, but is easier to state and so we give it first.

Theorem 2.15 *Boundedness Theorem*

Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function on $[a,b]$. Then f is bounded on $[a,b]$. That is, there is a real number M such that $|f(x)| \leq M$ for each $x \in [a,b]$.

This theorem states that continuous functions on closed bounded intervals must be bounded; that is, the image set of the closed bounded interval $[a,b]$ under f is a bounded set. Note that this is not necessarily the case if the domain is not bounded: for example, the function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = x$ is a continuous function, and the image set of $\mathbb{R}$ under f is again $\mathbb{R}$, an *un*bounded set.

Similarly, if f is a continuous function that has a domain which is an open interval, then f need not be bounded: consider $f\colon (0,1) \to \mathbb{R}$ given by $f(x) = \frac{1}{x}$.

Our next result tells us that a continuous function on a closed bounded interval is not only bounded but also attains its maximum and minimum values.

Theorem 2.16 *Extreme Value Theorem*

Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function on $[a,b]$. Then there are numbers $c, d \in [a,b]$ such that

$$f(c) \leq f(x) \leq f(d) \quad \text{for each } x \in [a,b].$$

Remarks

1. The theorem does not hold if either of the hypotheses (i.e. that f is continuous, and that the interval is closed and bounded) is dropped.
2. The number c is such that $f(c) \leq f(x)$ for all $x \in [a,b]$ – that is, $f(c)$ is the *minimum* value of the function in the interval. Similarly, $f(d)$ is the *maximum* value of the function in the interval.

Example 2.17

Consider $f\colon [-\pi,\pi] \to \mathbb{R}$ defined by $f(x) = e^x + 3\cos 2x$. Because f is continuous on $[-\pi,\pi]$ (see Worked Exercise 2.11), we can use the Extreme Value Theorem to deduce that there are points c and d in $[-\pi,\pi]$ such that $f(c) \leq f(x) \leq f(d)$ for all $x \in [-\pi,\pi]$. As Figure 2.11 indicates, c lies near -1.6 whilst $d = \pi$.

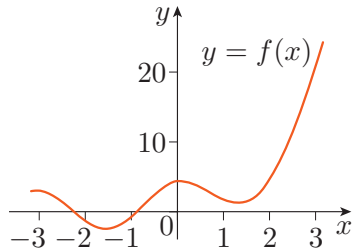


Figure 2.11 Graph of $f(x) = e^x + 3\cos 2x$ on $[-\pi,\pi]$

The Extreme Value Theorem is another example of an existence result: again, it does not tell us how to find the points whose existence it guarantees – we have to use other methods to do this.

Note that the Intermediate Value, Boundedness and Extreme Value Theorems hold for *all continuous functions* on *all closed bounded intervals*. We conclude that they are the result of an interaction between continuity and closed bounded intervals. When we look later at the properties of more general spaces, a generalisation of the idea of closed bounded intervals will give rise to the notion of 'compactness' – see Book F, Chapter 22.

Exercise 2.8

Let the function $f\colon [-1,1] - \{0\} \to \mathbb{R}$ be given by $f(x) = \frac{1}{x}$.

Explain why the Boundedness Theorem and the Extreme Value Theorem do not apply to this function.

3 Continuity on the plane

One of the main ways in which mathematics develops is by taking ideas from one situation and extending or generalising them into new ones. This process involves isolating some key aspects and ignoring others. We end this chapter by seeing whether we can use the definition of continuity for functions on the real line to come up with a definition of continuity for real-valued functions defined on the plane: functions with domain the plane, $\mathbb{R}^2$, and codomain $\mathbb{R}$. This will involve extending the ideas of distance, convergent sequences and continuity, which we developed in Sections 1 and 2, to make them appropriate for functions from $\mathbb{R}^2$ to $\mathbb{R}$. In the next chapter we will continue this process by first extending these ideas to spaces of higher dimensions and then considering even more general ideas of distance.

3.1 Functions from $\mathbb{R}^2$ to $\mathbb{R}$

In $\mathbb{R}^2$, we work with the familiar coordinate geometry of two dimensions. However, instead of using the common notation (x, y) for points in the plane, we use the notation (x_1, x_2) since this subscript notation is easy to extend to spaces of higher dimensions. We also use boldface letters such as $\mathbf{x}$ for the point (x_1, x_2), and $\mathbf{0}$ for the point $(0, 0)$. We refer to the point $\mathbf{0}$ as the **origin** of $\mathbb{R}^2$. For any two points $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$ in $\mathbb{R}^2$, we define

$$\mathbf{a} + \mathbf{b} = (a_1 + b_1, a_2 + b_2) \quad \text{and} \quad \mathbf{a} - \mathbf{b} = (a_1 - b_1, a_2 - b_2).$$

In specifying the rule of a function f with domain $\mathbb{R}^2$ we use the notation $f(\mathbf{x})$ or $f(x_1, x_2)$. Two examples of functions from $\mathbb{R}^2$ to $\mathbb{R}$ are given by

$$f(x_1, x_2) = 20 - (x_1^2 + x_2^2) \quad \text{and} \quad g(x_1, x_2) = \frac{x_1^2 x_2}{x_1^2 + x_2^2 + 1}.$$

Although, strictly speaking, $f(\mathbf{x}) = f((x_1, x_2))$, we omit one pair of brackets and write $f(x_1, x_2)$.

Just as we represent (the graphs of) functions from $\mathbb{R}$ to $\mathbb{R}$ by curves in the plane, the graphs of functions from $\mathbb{R}^2$ to $\mathbb{R}$ are represented by *surfaces* in three-dimensional space. Above each point (x_1, x_2) in the domain $\mathbb{R}^2$, which we draw as a horizontal plane, we mark a point whose third coordinate x_3 is the function value, $f(x_1, x_2)$. Thus, using the examples given, above the point $(3, 2)$ we mark the point $f(3, 2) = 20 - 13 = 7$. As (x_1, x_2) ranges over different values in the domain $\mathbb{R}^2$, the values of $x_3 = f(x_1, x_2)$ define a surface in space. The graphs of the functions f and g are shown as surfaces in three-dimensional space in Figure 3.1.

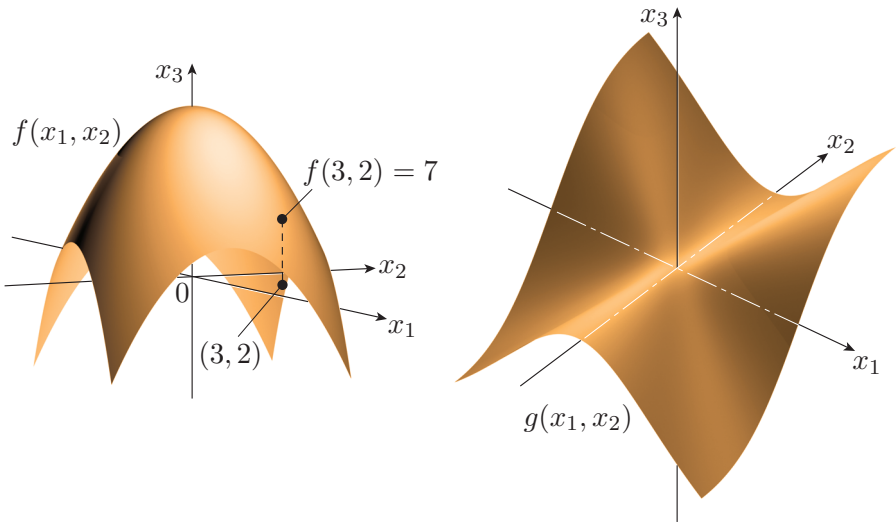


Figure 3.1 Graphs of $f, g\colon \mathbb{R}^2 \to \mathbb{R}$ given by $f(x_1, x_2) = 20 - (x_1^2 + x_2^2)$ and $g(x_1, x_2) = \dfrac{x_1^2 x_2}{x_1^2 + x_2^2 + 1}$

As a first step in considering the continuity of functions on $\mathbb{R}^2$, we follow the development in Sections 1 and 2 and look at sequences in $\mathbb{R}^2$.

3.2 Sequences in $\mathbb{R}^2$

In Section 2 we defined continuity for functions from $\mathbb{R}$ to $\mathbb{R}$ in terms of the convergence of appropriate sequences of real numbers. We now discuss sequences of points in the plane $\mathbb{R}^2$, and show how to define convergence for these. This will allow us to define the continuity of functions from $\mathbb{R}^2$ to $\mathbb{R}$ (in the next subsection).

Definition 3.1 *Sequence in* $\mathbb{R}^2$

A **sequence in** $\mathbb{R}^2$ is an unending ordered list of points in $\mathbb{R}^2$:

$$\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \ldots.$$

The point $\mathbf{a}_n = (a_{1,n}, a_{2,n})$ is the ***n*th term** of the sequence, and the whole sequence is denoted by $(\mathbf{a}_n)$, $(\mathbf{a}_n)_{n=1}^{\infty}$ or $(\mathbf{a}_n)_{n\in\mathbb{N}}$.

Remarks

1. Each term in the sequence is a point in the two-dimensional plane $\mathbb{R}^2$. This explains the double subscript: the point $\mathbf{a}_n$ has first coordinate $a_{1,n}$ and second coordinate $a_{2,n}$. So the nth term $\mathbf{a}_n$ is in $\mathbb{R}^2$, whereas its coordinates $a_{1,n}$ and $a_{2,n}$ are real numbers.
2. We could also write the sequence $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \ldots$ as

$$(a_{1,1}, a_{2,1}),\ (a_{1,2}, a_{2,2}),\ (a_{1,3}, a_{2,3}),\ \ldots.$$

Example 3.2

Examples of sequences in the plane include the sequence

$$(\mathbf{a}_n) \text{ where } \mathbf{a}_n = (\cos n, \sin n),$$

consisting of certain points on the unit circle, and the sequence

$$(\mathbf{b}_n) \text{ where } \mathbf{b}_n = \left(\tfrac{1}{n}, n\right),$$

consisting of certain points on the graph of $f\colon (0,\infty) \to \mathbb{R}$ given by $f(x) = \frac{1}{x}$ (see Figure 3.2).

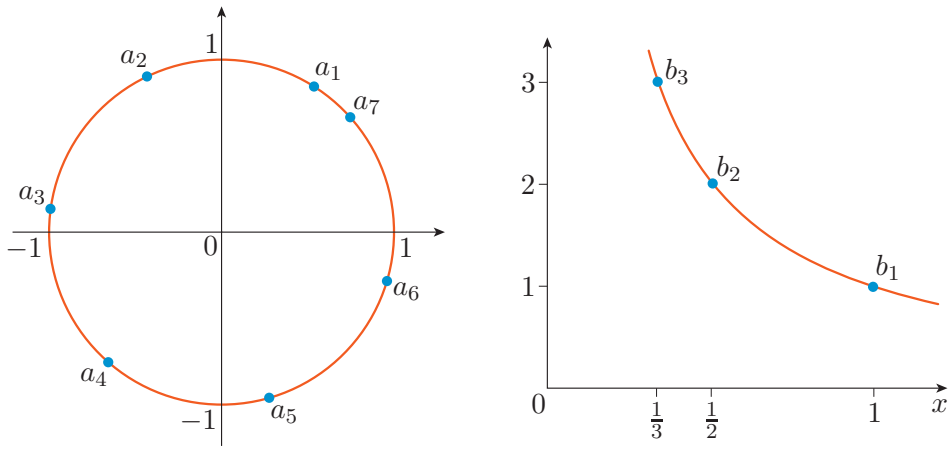


Figure 3.2 Examples of sequences in the plane

For sequences in $\mathbb{R}$, we have the notion of convergence: recall that a sequence (a_n) of real numbers is convergent (with limit l) if for each $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that $|a_n - l| < \varepsilon$ whenever $n > N$. Notice that $|a_n - l|$ measures the distance between the two real numbers a_n and l, and that if we replaced this term with the distance between pairs of (appropriate) points in the plane, we would (potentially) have a definition of convergence for sequences of points in the plane. This motivates the following brief digression.

Distance in the plane

We saw in Section 1 that on the real line $\mathbb{R}$, the distance between two points a and b is given by $|b - a|$. In $\mathbb{R}^2$ we want the distance between two points to be defined in the usual way: for example, as Figure 3.3 illustrates, the distance between the points $(3, -1)$ and $(7, 2)$ is

$$\sqrt{(7-3)^2 + (2-(-1))^2} = \sqrt{4^2 + 3^2} = 5.$$

This distance, which is calculated by Pythagoras's Theorem, is called the *Euclidean distance* in $\mathbb{R}^2$.

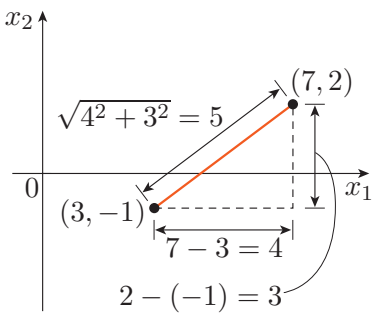


Figure 3.3 Distance between $(3, -1)$ and $(7, 2)$

Definition 3.3 *Euclidean distance in* $\mathbb{R}^2$

The **Euclidean distance** between points $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$ in $\mathbb{R}^2$ is given by the formula

$$d^{(2)}(\mathbf{a}, \mathbf{b}) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}.$$

Putting $\mathbf{b} = \mathbf{0} = (0, 0)$ in the definition of $d^{(2)}(\mathbf{a}, \mathbf{b})$ gives

$$d^{(2)}(\mathbf{a}, \mathbf{0}) = \sqrt{a_1^2 + a_2^2},$$

which is the distance between $\mathbf{a}$ and the origin, $\mathbf{0}$.

If you think of $\mathbf{a}$ as a vector, then $d^{(2)}(\mathbf{a}, \mathbf{0})$ is the *magnitude* of $\mathbf{a}$, often denoted by $||\mathbf{a}||$ or $|\mathbf{a}|$.

When calculating the distance between two points in the plane, it is frequently convenient to use the fact that

$$d^{(2)}(\mathbf{a}, \mathbf{b}) = d^{(2)}(\mathbf{a} - \mathbf{b}, \mathbf{0}) = d^{(2)}(\mathbf{b} - \mathbf{a}, \mathbf{0}), \quad \text{for } \mathbf{a}, \mathbf{b} \in \mathbb{R}^2.$$

Exercise 3.1

Prove that, for each $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$,

$$d^{(2)}(\mathbf{a}, \mathbf{b}) = d^{(2)}(\mathbf{a} - \mathbf{b}, \mathbf{0}) = d^{(2)}(\mathbf{b} - \mathbf{a}, \mathbf{0}).$$

These relationships have a geometric interpretation. When the line segment between $\mathbf{a}$ and $\mathbf{b}$ is translated so that one of its ends is at the origin, the distance between its endpoints is unchanged, as Figure 3.4 illustrates.

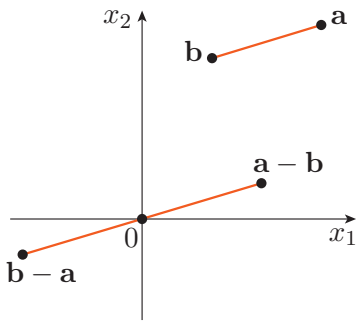


Figure 3.4 Translating a line segment leaves the distance between its endpoints unchanged

Convergence of sequences in the plane

Now that we know how to measure distances in the plane, we can define convergence of planar sequences.

Recall the following result from Definition 1.17.

For a real sequence (a_n), $a_n \to l$ as $n \to \infty$ if, and only if, for each $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that

$$|a_n - l| < \varepsilon \quad \text{whenever } n > N.$$

Thus if we replace $|a_n - l|$ with $d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell})$ for $\mathbf{a}_n$ and $\boldsymbol{\ell}$ points in $\mathbb{R}^2$, then we have a possible definition for convergence of sequences in the plane: we could say that a sequence $(\mathbf{a}_n)$ in $\mathbb{R}^2$ converges to a point $\boldsymbol{\ell} \in \mathbb{R}^2$ if for each $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that $d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell}) < \varepsilon$ whenever $n > N$. But now consider the sequence (b_n) given by $b_n = d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell})$ for each $n \in \mathbb{N}$. This is a sequence of (non-negative) real numbers that converges to zero as $n \to \infty$ – that is, it is a null sequence – and so this is an alternative way of saying the same thing.

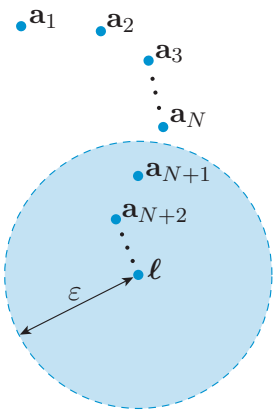


Figure 3.5 Convergence of $(\mathbf{a}_n)$

See Exercise 3.1.

Definition 3.4 *Convergence in* $\mathbb{R}^2$

A sequence $(\mathbf{a}_n)$ in $\mathbb{R}^2$ **converges** to $\boldsymbol{\ell} \in \mathbb{R}^2$ if $\left(d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell})\right)$ is a (real) null sequence.

We say that $\boldsymbol{\ell}$ is the **limit** of the sequence $(\mathbf{a}_n)$ and write $\mathbf{a}_n \to \boldsymbol{\ell}$ as $n \to \infty$, $\lim_{n\to\infty} \mathbf{a}_n = \boldsymbol{\ell}$ or simply $\mathbf{a}_n \to \boldsymbol{\ell}$.

A sequence in the plane that does not converge to any point in $\mathbb{R}^2$ is said to be **divergent**.

Figure 3.5 illustrates a sequence $(\mathbf{a}_n)$ converging to $\boldsymbol{\ell}$, with $d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell}) < \varepsilon$ for all $n > N$.

Remarks

1. Since $d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell}) = d^{(2)}(\mathbf{a}_n - \boldsymbol{\ell}, \mathbf{0})$, the statement '$(\mathbf{a}_n)$ converges to $\boldsymbol{\ell}$' is equivalent to '$(\mathbf{a}_n - \boldsymbol{\ell})$ is a sequence that converges to $\mathbf{0}$'.
2. By use of the combination rules for null sequences (in $\mathbb{R}$), we see that

   $$d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell}) \to 0 \text{ if, and only if, } d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell})^2 \to 0$$

   and so to show that $\mathbf{a}_n \to \boldsymbol{\ell}$ as $n \to \infty$, it is enough to show that

   $$(a_{1,n} - l_1)^2 + (a_{2,n} - l_2)^2 \to 0 \text{ as } n \to \infty.$$

Worked Exercise 3.5

Let $(\mathbf{a}_n)$ be the sequence in the plane defined by $\mathbf{a}_n = \left(\frac{1}{n^2}, 5 + \frac{1}{n}\right)$. Show that $(\mathbf{a}_n)$ converges to $(0, 5)$.

Solution

It is enough to show that $(d^{(2)}(\mathbf{a}_n, (0,5))^2)$ is a (real) null sequence.

For $n \in \mathbb{N}$, we have

$$\begin{aligned} d^{(2)}\left(\left(\tfrac{1}{n^2}, 5 + \tfrac{1}{n}\right), (0,5)\right)^2 &= \left(\tfrac{1}{n^2}\right)^2 + \left(\tfrac{1}{n}\right)^2 \\ &= \tfrac{1}{n^4} + \tfrac{1}{n^2}, \end{aligned}$$

a sum of basic null sequences. We conclude (from the combination rules for null sequences) that $(d^{(2)}(\mathbf{a}_n, (0,5))^2)$ is a (real) null sequence.

Hence $\mathbf{a}_n \to (0, 5)$ as $n \to \infty$.

Worked Exercise 3.6

Let $(\mathbf{b}_n)$ be the sequence in the plane defined by $\mathbf{b}_n = (\frac{1}{n}, n)$. Show that $(\mathbf{b}_n)$ is a divergent sequence.

Solution

Let $\boldsymbol{\ell} = (l_1, l_2)$ be a possible point in the plane that $(\boldsymbol{\ell}_n)$ converges to and notice that

$$d^{(2)}(\mathbf{b}_n, \boldsymbol{\ell})^2 = (\tfrac{1}{n} - \ell_1)^2 + (n - \ell_2)^2 \geq (n - \ell_2)^2.$$

But if $N = \text{int}(|\ell_2|) + 1$, then for $n > N$, $(n - \ell_2)^2 > 1$ and so $d^{(2)}(\mathbf{b}_n, \boldsymbol{\ell})^2$ is not a null sequence.

Hence $(\mathbf{b}_n)$ is a divergent sequence.

Exercise 3.2

Let $(\mathbf{b}_n)$ be the sequence defined by $\mathbf{b}_n = \left(\frac{3}{n} + 1, 4 - \frac{2}{n^3}\right)$. Prove that $(\mathbf{b}_n)$ converges, and find its limit.

In the exercises above, you may notice a connection between the convergence of a sequence of points in the plane and the convergence of the individual coordinates of the terms in the sequence. For example, in Worked Exercise 3.5 the sequence given by $\mathbf{a}_n = \left(\frac{1}{n^2}, 5 + \frac{1}{n}\right)$ converges to $(0, 5)$, while the coordinate sequences $\left(\frac{1}{n^2}\right)$ and $\left(5 + \frac{1}{n}\right)$ converge to 0 and 5 respectively. The next theorem shows that this holds true in general.

Theorem 3.7

Let $(\mathbf{a}_n) = ((a_{1,n}, a_{2,n}))$ be a sequence in $\mathbb{R}^2$ and let $\boldsymbol{\ell} = (\ell_1, \ell_2) \in \mathbb{R}^2$. Then

$\mathbf{a}_n \to \boldsymbol{\ell}$ if, and only if, both $a_{1,n} \to \ell_1$ and $a_{2,n} \to \ell_2$ as $n \to \infty$.

Proof Let us first suppose that $(\mathbf{a}_n)$ converges to $\boldsymbol{\ell}$ and prove that $a_{1,n} \to \ell_1$ and $a_{2,n} \to \ell_2$. This means that we need to show that the real sequences $(a_{1,n} - \ell_1)_{n\in\mathbb{N}}$ and $(a_{2,n} - \ell_2)_{n\in\mathbb{N}}$ are null. We estimate

$$0 \leq |a_{1,n} - \ell_1| \leq \sqrt{(a_{1,n} - \ell_1)^2 + (a_{2,n} - \ell_2)^2} = d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell}).$$

We are assuming that $(\mathbf{a}_n)$ converges to $\boldsymbol{\ell}$, and so the real sequence $(d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell}))_{n\in\mathbb{N}}$ must be null. Thus it follows from the Squeeze Rule that $(a_{1,n} - \ell_1)_{n\in\mathbb{N}}$ is null.

Hence $a_{1,n} \to \ell_1$. The proof that $a_{2,n} \to \ell_2$ as $n \to \infty$ is similar.

To complete the proof, we now suppose that $a_{1,n} \to \ell_1$ and $a_{2,n} \to \ell_2$, and show that this implies $\mathbf{a}_n \to \boldsymbol{\ell}$ as $n \to \infty$.

So, given that $(a_{1,n} - \ell_1)$ and $(a_{2,n} - \ell_2)$ are null sequences, we use the Product and Sum Rules to deduce that $\left((a_{1,n} - \ell_1)^2 + (a_{2,n} - \ell_2)^2\right)$ is a null sequence of non-negative real numbers. Then the Power Rule for null sequences (see Theorem 1.14) gives that the sequence

$$(d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell})) = \left(\sqrt{(a_{1,n} - \ell_1)^2 + (a_{2,n} - \ell_2)^2}\right)$$

is null. We conclude that $\mathbf{a}_n$ converges to $\boldsymbol{\ell}$ as $n \to \infty$, as required. ■

Remarks

1. Thus a sequence $(\mathbf{a}_n)$ in the plane is divergent if, and only if, at least one of the component sequences $(a_{1,n})$ and $(a_{2,n})$ is divergent.
2. This result enables us to prove combination rules for sequences in the plane based upon those for real sequences.

Exercise 3.3

(a) Let $(\mathbf{a}_n)$ be the sequence defined by $\mathbf{a}_n = \left(\frac{n^2}{3^n}, \frac{3n! + 2^n}{n^{50} + n!}\right)$. Show that $(\mathbf{a}_n)$ converges, and find its limit.

(b) Suppose that $(\mathbf{a}_n)$ and $(\mathbf{b}_n)$ are sequences in the plane that converge to $\boldsymbol{\ell}$ and $\mathbf{m}$ respectively, as $n \to \infty$. Show that the sequence $(\mathbf{c}_n)$ given by $\mathbf{c}_n = \mathbf{a}_n + \mathbf{b}_n$ converges to $\boldsymbol{\ell} + \mathbf{m}$ as $n \to \infty$.

We could, in a similar way, prove other combination rules for sequences in the plane: we avoid the temptation to do so.

3.3 Continuity on the plane

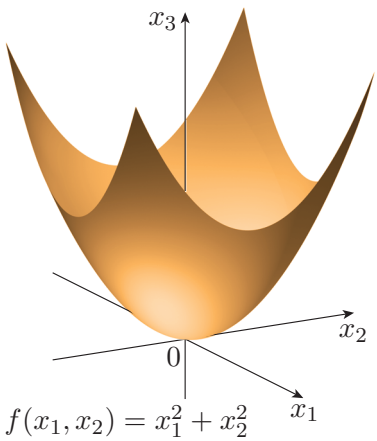


Figure 3.6 Graph of $f(x_1, x_2) = x_1^2 + x_2^2$

In the final part of this section (and chapter), we pull together what we have learned to define continuity for functions from the plane to the real line.

Consider the graph of the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x_1, x_2) = x_1^2 + x_2^2,$$

shown in Figure 3.6. There are no obvious jumps or gaps, and no regions where the function appears to behave oddly.

It certainly seems reasonable to expect that for any reasonable definition of continuity, this function should be continuous on $\mathbb{R}^2$. However, the definition of continuity given in Section 2 applies only to functions from $\mathbb{R}$ to $\mathbb{R}$. To prove continuity for functions from $\mathbb{R}^2$ to $\mathbb{R}$, such as this one, we need a new definition, which should be related in a natural way to the one for functions from $\mathbb{R}$ to $\mathbb{R}$. To show how this can be done, we first give a reminder of the definition of continuity from $\mathbb{R}$ to $\mathbb{R}$.

Continuity from $\mathbb{R}$ to $\mathbb{R}$

Let $A \subseteq \mathbb{R}$. A function $f\colon A \to \mathbb{R}$ is **continuous** at $a \in A$ if:

for each sequence (x_n) in A for which $x_n \to a$ as $n \to \infty$,
the sequence $(f(x_n))$ converges to $f(a)$.

We need to change this definition to refer to the fact that the domain of f is a subset of $\mathbb{R}^2$. The changes are fairly straightforward: A becomes a subset of $\mathbb{R}^2$; the point a in $\mathbb{R}$ becomes a point $\mathbf{a}$ in $\mathbb{R}^2$; and in place of real sequences (x_n) we consider sequences $(\mathbf{x}_n)$ in the plane. These changes lead to the following definition.

Definition 3.8 *Continuity at a point in the plane*

Let $A \subseteq \mathbb{R}^2$ and let $f\colon A \to \mathbb{R}$ be a function.

Then f is **continuous** at $\mathbf{a} \in A$ if:

whenever $(\mathbf{x}_n)$ is a sequence in A for which $\mathbf{x}_n \to \mathbf{a}$ as $n \to \infty$,
then the (real) sequence $(f(\mathbf{x}_n))$ converges to $f(\mathbf{a})$.

If f does not satisfy this condition at $\mathbf{a} \in A$ – that is, there is a sequence $(\mathbf{x}_n)$ in A for which $\mathbf{x}_n \to \mathbf{a}$ as $n \to \infty$ but $f(\mathbf{x}_n)$ does not converge to $f(\mathbf{a})$ – then we say that f is **discontinuous** at $\mathbf{a}$.

To show that f is continuous at $\mathbf{a}$, we must show that $(f(\mathbf{x}_n))$ converges to $f(\mathbf{a})$ for *every possible choice* of a sequence $(\mathbf{x}_n)$ in A that converges to $\mathbf{a}$.

We extend the definition of continuity at a point to define continuity on a subset of the plane.

Definition 3.9 *Continuity on a subset of the plane*

Let $S \subseteq A \subseteq \mathbb{R}^2$ and let $f\colon A \to \mathbb{R}$ be a function.

We say that f is **continuous** on S when f is continuous at each point in S.

We say that f is **discontinuous** on S if f is discontinuous at at least one point in S.

Our first examples of continuous functions from the plane to the real line are very simple but will prove useful later.

Definition 3.10 *Projection functions in the plane*

The two functions $p_1\colon \mathbb{R}^2 \to \mathbb{R}$ and $p_2\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$p_1(x_1, x_2) = x_1, \quad p_2(x_1, x_2) = x_2,$$

are known as **projection functions**.

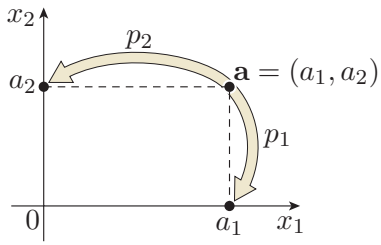


Figure 3.7 The projection functions

These functions *project* each point $\mathbf{a} = (a_1, a_2)$ in $\mathbb{R}^2$ onto the coordinate axes, as Figure 3.7 illustrates.

Worked Exercise 3.11

Prove that the projection functions p_1 and p_2 are continuous on $\mathbb{R}^2$.

Solution

We give the proof for p_1. The proof for p_2 is similar.

Let $\mathbf{a} = (a_1, a_2) \in \mathbb{R}^2$ be a general point in the plane, and let $(\mathbf{x}_n) = ((x_{1,n}, x_{2,n}))$ be a sequence in $\mathbb{R}^2$ that converges to $\mathbf{a}$. We need to show that $(p_1(\mathbf{x}_n))$ converges to $p_1(\mathbf{a})$.

We have that $p_1(\mathbf{x}_n) = x_{1,n}$ and $p_1(\mathbf{a}) = a_1$.

Since we are assuming that $\mathbf{x}_n \to \mathbf{a}$, we deduce from Theorem 3.7 that $x_{1,n} \to a_1$.

Hence the sequence $(p_1(\mathbf{x}_n))_{n\in\mathbb{N}} = (x_{1,n})_{n\in\mathbb{N}}$ converges to $p_1(\mathbf{a}) = a_1$, and therefore p_1 is continuous at $\mathbf{a}$.

Since $\mathbf{a}$ is an arbitrary point, p_1 is continuous on $\mathbb{R}^2$.

To show that a function from $\mathbb{R}^2$ to $\mathbb{R}$ is discontinuous, we need to find *just one sequence* $\mathbf{x}_n \to \mathbf{a}$ such that $(f(\mathbf{x}_n))$ does not converge to $f(\mathbf{a})$.

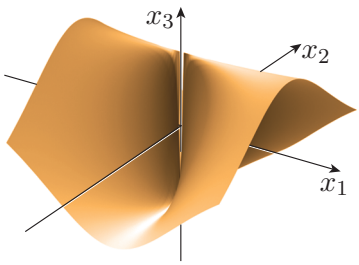


Figure 3.8 Graph of function for Worked Exercise 3.12

Worked Exercise 3.12

Show that the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(\mathbf{x}) = \begin{cases} \dfrac{x_1^2 - x_2^2}{x_1^2 + x_2^2}, & \text{if } \mathbf{x} \neq \mathbf{0}, \\ 0, & \text{if } \mathbf{x} = \mathbf{0}, \end{cases}$$

is discontinuous at $\mathbf{0}$. (The graph of the function is shown in Figure 3.8.)

Solution

Look along the positive x_1-axis, where $x_2 = 0$. There, $f(\mathbf{x}) = f(x_1, 0) = 1$, except at the origin $\mathbf{0}$, where $f(\mathbf{0}) = 0$.

In order to show discontinuity at $\mathbf{0}$, we need only to find a single sequence $\mathbf{x}_n \to \mathbf{0}$ such that $f(\mathbf{x}_n) \not\to f(\mathbf{0}) = 0$ as $n \to \infty$.

Using the behaviour of f along the x_1-axis as our guide, we choose $\mathbf{x}_n = (\frac{1}{n}, 0)$ for $n \in \mathbb{N}$. Then

$$d^{(2)}(\mathbf{x}_n, \mathbf{0}) = \tfrac{1}{n} \to 0 \text{ as } n \to \infty,$$

and so $\mathbf{x}_n \to \mathbf{0}$.

However,

$$f(\mathbf{x}_n) = f(\tfrac{1}{n}, 0) = 1, \quad \text{for each } n \in \mathbb{N},$$

and so $(f(\mathbf{x}_n))$ is just the constant sequence (1), which does not converge to $f(\mathbf{0}) = 0$.

Hence f is discontinuous at $\mathbf{0}$.

Any sequence given by $\mathbf{x}_n = (x_{1,n}, 0)$ with $x_{1,n} \neq 0$ and $x_{1,n} \to 0$ would do.

Along the line $x_1 = x_2$, $f(\mathbf{x})$ is always 0. Although this alone might suggest that f is continuous at $\mathbf{0}$, this is not the case. To be continuous at $\mathbf{0}$, it is necessary that $f(\mathbf{x})$ must approach the value at $f(\mathbf{0})$ *however* $\mathbf{x}$ *approaches* $\mathbf{0}$.

Exercise 3.4

Show that the function $g\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$g(x_1, x_2) = \begin{cases} 0, & \text{if } x_2 \geq x_1, \\ 1, & \text{if } x_2 < x_1, \end{cases}$$

is discontinuous at $(2, 2)$.

The next worked exercise involves a more complicated function and the solution requires some extra effort.

Worked Exercise 3.13

Prove that the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x_1, x_2) = x_1^2 + x_2^2$$

is continuous on $\mathbb{R}^2$. (The graph of the function is shown in Figure 3.9.)

Solution

Let $\mathbf{a} \in \mathbb{R}^2$ and let $(\mathbf{x}_n) = ((x_{1,n}, x_{2,n}))$ be a sequence that converges to $\mathbf{a}$.

We prove that $(f(\mathbf{x}_n))$ converges to $f(\mathbf{a})$ by showing that the sequence of real numbers $(f(\mathbf{x}_n) - f(\mathbf{a}))_{n\in\mathbb{N}}$ is null.

First, we seek an upper bound for $|f(\mathbf{x}_n) - f(\mathbf{a})|$ in terms of $d^{(2)}(\mathbf{x}_n, \mathbf{a})$:

$$\begin{aligned} |f(\mathbf{x}_n) - f(\mathbf{a})| &= |(x_{1,n}^2 + x_{2,n}^2) - (a_1^2 + a_2^2)| \\ &= |(x_{1,n}^2 - a_1^2) + (x_{2,n}^2 - a_2^2)| \\ &\leq |x_{1,n}^2 - a_1^2| + |x_{2,n}^2 - a_2^2| \text{ (by the Triangle Inequality)} \\ &= |x_{1,n} - a_1|\,|x_{1,n} + a_1| + |x_{2,n} - a_2|\,|x_{2,n} + a_2|. \quad (1) \end{aligned}$$

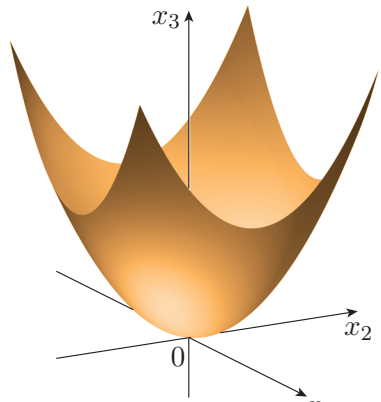


Figure 3.9 Graph of function for Worked Exercise 3.13

But

$$|x_{1,n} - a_1| = \sqrt{(x_{1,n} - a_1)^2}$$
$$\leq \sqrt{(x_{1,n} - a_1)^2 + (x_{2,n} - a_2)^2} = d^{(2)}(\mathbf{x}_n, \mathbf{a}), \quad (2)$$

and, similarly,

$$|x_{2,n} - a_2| \leq d^{(2)}(\mathbf{x}_n, \mathbf{a}). \quad (3)$$

Hence, from equations (1), (2) and (3),

$$|f(\mathbf{x}_n) - f(\mathbf{a})| \leq (|x_{1,n} + a_1| + |x_{2,n} + a_2|)d^{(2)}(\mathbf{x}_n, \mathbf{a}). \quad (4)$$

It still remains to find upper bounds for $|x_{1,n} + a_1|$ and $|x_{2,n} + a_2|$. To do this for $|x_{1,n} + a_1|$, we rewrite it as $|(x_{1,n} - a_1) + 2a_1|$ and use the Triangle Inequality and equation (2) to obtain

$$|x_{1,n} + a_1| = |(x_{1,n} - a_1) + 2a_1| \leq |x_{1,n} - a_1| + 2|a_1|$$
$$\leq d^{(2)}(\mathbf{x}_n, \mathbf{a}) + 2|a_1|. \quad (5)$$

Similarly, for $|x_{2,n} + a_2|$,

$$|x_{2,n} + a_2| \leq d^{(2)}(\mathbf{x}_n, \mathbf{a}) + 2|a_2|. \quad (6)$$

Hence, from equations (4), (5) and (6),

$$0 \leq |f(\mathbf{x}_n) - f(\mathbf{a})| \leq \left(2d^{(2)}(\mathbf{x}_n, \mathbf{a}) + 2|a_1| + 2|a_2|\right) d^{(2)}(\mathbf{x}_n, \mathbf{a}). \quad (7)$$

Since $\mathbf{x}_n \to \mathbf{a}$, the sequence $\left(d^{(2)}(\mathbf{x}_n, \mathbf{a})\right)_{n\in\mathbb{N}}$ is null and so the sequence $\left(2d^{(2)}(\mathbf{x}_n, \mathbf{a}) + 2|a_1| + 2|a_2|\right)$ converges to $2|a_1| + 2|a_2|$ by the Sum Rule. The Product Rule now gives that

$$\left(2d^{(2)}(\mathbf{x}_n, \mathbf{a}) + 2|a_1| + 2|a_2|\right) d^{(2)}(\mathbf{x}_n, \mathbf{a}) \to (2|a_1| + 2|a_2|) \times 0 = 0.$$

That is, the right-hand side of equation (7) is a null sequence. By applying the Squeeze Rule to equation (7), we deduce that $(f(\mathbf{x}_n) - f(\mathbf{a})) \to 0$ as $n \to \infty$.

Therefore $f(\mathbf{x}_n) \to f(\mathbf{a})$, and so f is continuous at $\mathbf{a}$.

Since $\mathbf{a}$ is an arbitrary point, f is continuous on $\mathbb{R}^2$.

As Worked Exercise 3.13 illustrates, it would be tedious if we had to work directly from the definition of continuity whenever we had to prove that a function from $\mathbb{R}^2$ to $\mathbb{R}$ is continuous. Fortunately, as with functions from $\mathbb{R} \to \mathbb{R}$, there are combination rules that allow us to combine continuous functions to form new continuous functions. We state them without proof.

Combination rules for continuous functions from $\mathbb{R}^2$ to $\mathbb{R}$

Let $A \subseteq \mathbb{R}^2$ and let $f\colon A \to \mathbb{R}$ and $g\colon A \to \mathbb{R}$ be functions that are continuous at $\mathbf{a} \in A$.

Then the following functions are continuous at $\mathbf{a}$:

Sum Rule	$f + g$, defined by $(f+g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$
Multiple Rule	λf where $\lambda \in \mathbb{R}$, defined by $(\lambda f)(\mathbf{x}) = \lambda \times f(\mathbf{x})$
Product Rule	fg, defined by $(fg)(\mathbf{x}) = f(\mathbf{x})g(\mathbf{x})$
Quotient Rule	f/g, defined by $(f/g)(\mathbf{x}) = f(\mathbf{x})/g(\mathbf{x})$, provided $g(\mathbf{a}) \neq 0$.

These combination rules for continuous functions from $\mathbb{R}^2$ to $\mathbb{R}$ follow from the combination rules for convergent real sequences.

Worked Exercise 3.14

Let $\mathbf{u} = (u_1, u_2)$ be a fixed point in $\mathbb{R}^2$. Use the combination rules to prove that the function $g\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$g(\mathbf{x}) = u_1x_1 + u_2x_2$$

is continuous on $\mathbb{R}^2$.

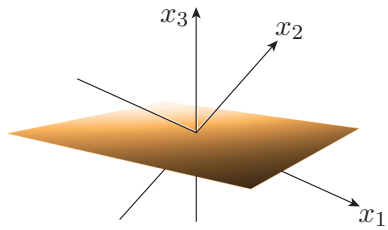
Note that the graph of g is a plane through the origin, as Figure 3.10 illustrates for $\mathbf{u} = (2, -3)$. It is intuitively clear that g is continuous everywhere, but this still needs proof.

Solution

The key to proving that this function is continuous is to recognise that g can be written in terms of the projection functions p_1 and p_2. In fact,

$$g(\mathbf{x}) = u_1p_1(\mathbf{x}) + u_2p_2(\mathbf{x}).$$

Since both p_1 and p_2 are continuous on $\mathbb{R}^2$ and since $\mathbf{u} = (u_1, u_2)$ is fixed, the Multiple Rule guarantees that u_1p_1 and u_2p_2 are both continuous functions on $\mathbb{R}^2$. We can now apply the Sum Rule to u_1p_1 and u_2p_2 to deduce that $g = u_1p_1 + u_2p_2$ is continuous on $\mathbb{R}^2$.



$g(x_1, x_2) = 2x_1 - 3x_2$

Figure 3.10 Graph of $g(\mathbf{x})$ for $\mathbf{u} = (2, -3)$

The result of Worked Exercise 3.14 applies for any $\mathbf{u}$; so it tells us, for example, that the function $f(\mathbf{x}) = 2x_1 - 3x_2$, which is $g(\mathbf{x})$ with $\mathbf{u} = (2, -3)$, is continuous on $\mathbb{R}^2$.

A similar argument works for other functions: for example, an alternative way to verify the continuity of the function f in Worked Exercise 3.13 is to notice that if $f(x_1, x_2) = x_1^2 + x_2^2$, then $f(\mathbf{x}) = (p_1(\mathbf{x}))^2 + (p_2(\mathbf{x}))^2$ and so, since (by Worked Exercise 3.11) p_1 and p_2 are continuous on $\mathbb{R}^2$, we can use the Product Rule (applied to $p_1 \times p_1$ and $p_2 \times p_2$) followed by the Sum Rule to deduce that f is continuous on $\mathbb{R}^2$.

There is also a Restriction Rule for functions from $\mathbb{R}^2$ to $\mathbb{R}$, which allows us to restrict the domain of a continuous function to a smaller set and still have a continuous function.

See Definition 2.7 for the definition of the restricted function.

Proposition 3.15 *Restriction Rule*

Let $A \subseteq \mathbb{R}^2$, let $f\colon A \to \mathbb{R}$ and let $B \subseteq A$. If f is continuous at $\mathbf{b} \in B$, then the restricted function $f|_B\colon B \to \mathbb{R}$ is continuous at $\mathbf{b}$.

In particular, if f is continuous on A, then $f|_B$ is continuous on B.

Another useful rule is the one that allows us to compose continuous functions from $\mathbb{R}^2$ to $\mathbb{R}$ with continuous functions from $\mathbb{R}$ to $\mathbb{R}$. Again we state it without proof.

Proposition 3.16 *Composition Rule*

Let $A \subseteq \mathbb{R}^2$ and suppose that $f\colon A \to \mathbb{R}$ is continuous on A. Let $f(A) \subseteq B \subseteq \mathbb{R}$ and suppose $g\colon B \to \mathbb{R}$ is continuous on B. Then the composed function $g \circ f\colon A \to \mathbb{R}$ is continuous on A.

Notice that we are *not* composing two functions from $\mathbb{R}^2$ to $\mathbb{R}$. Since the codomain of f is $\mathbb{R}$, and not $\mathbb{R}^2$, the domain of g must be a subset of $\mathbb{R}$. Thus g maps from some subset of $\mathbb{R}$ to $\mathbb{R}$.

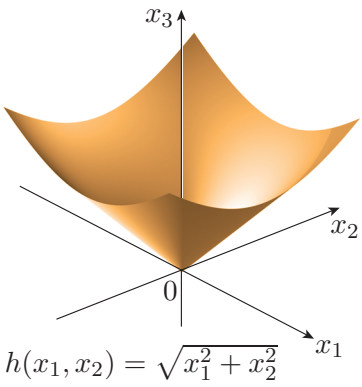


Figure 3.11 Graph of $h(\mathbf{x})$ for Worked Exercise 3.17

Worked Exercise 3.17

Prove that the function $h\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$h(\mathbf{x}) = \sqrt{x_1^2 + x_2^2}$$

is continuous on $\mathbb{R}^2$. (The graph of the function is shown in Figure 3.11.)

Solution

We see that h consists of composing the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by $f(x_1, x_2) = x_1^2 + x_2^2$ with the square-root function. We have already shown that f is continuous on $\mathbb{R}^2$ (Worked Exercise 3.13) and we know that the square-root function is a basic continuous function from $[0, \infty)$ to $\mathbb{R}$. Furthermore, since $f(x_1, x_2) = x_1^2 + x_2^2 \geq 0$ for all real numbers, the image set of f is included in the domain $[0, \infty)$ of the square-root function. Hence the conditions of the Composition Rule are satisfied, so we conclude that h is continuous on $\mathbb{R}^2$.

Exercise 3.5

Use the projection functions and the rules for combining continuous functions to prove that the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(\mathbf{x}) = \sin(x_1 + x_2)$$

is continuous on $\mathbb{R}^2$.

Exercise 3.6

Use the projection functions and the rules for combining continuous functions to prove that the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(\mathbf{x}) = e^{-x_1 x_2}$$

is continuous on $\mathbb{R}^2$.

3.4 Properties of distance

Because we wish to generalise the idea of distance to higher dimensions, and to other contexts, we finish this chapter by highlighting certain properties of the usual Euclidean distance.

So far, we have met two ways of defining distance: the distance between points a and b on the real line given by $|b - a|$, and the distance between points $\mathbf{a}$ and $\mathbf{b}$ in the plane given by

$$d^{(2)}(\mathbf{a}, \mathbf{b}) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}.$$

Another definition of distance with which you may be familiar is the great circle distance between points on the surface of a sphere. You will meet other kinds of distance in Chapter 14.

Both of these notions of distance can be thought of as functions of two variables. On the real line, we associate with each pair of real numbers a and b a single non-negative real number $|b - a|$. Similarly, on the plane, we associate with each pair of points (a_1, a_2) and (b_1, b_2) the single non-negative real number

$$\sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}.$$

Both functions are referred to as *Euclidean distance*, on $\mathbb{R}$ and on $\mathbb{R}^2$ respectively.

Each of these distance functions has three significant properties, the first two of which may seem almost trivial. We draw attention to these three properties because they are the key properties that we will use to define the general notion of a distance function. We concentrate on describing these properties for the Euclidean distance function $d^{(2)}$ on $\mathbb{R}^2$.

The first property is that the distance between two *distinct* points is always positive, and the distance between two points that coincide is 0. This property, which we label (M1), is stated formally as follows.

M stands for *metric* – of which you will see more in the next chapter.

(M1) For each $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$, we have $d^{(2)}(\mathbf{a}, \mathbf{b}) \geq 0$, with $d^{(2)}(\mathbf{a}, \mathbf{b}) = 0$ if, and only if, $\mathbf{a} = \mathbf{b}$.

The second property is that it does not matter in which direction you measure the distance between two points: the distance from $\mathbf{a}$ to $\mathbf{b}$ is the same as the distance from $\mathbf{b}$ to $\mathbf{a}$. We say that $d^{(2)}$ is *symmetric*.

(M2) For each $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$, we have $d^{(2)}(\mathbf{a}, \mathbf{b}) = d^{(2)}(\mathbf{b}, \mathbf{a})$.

Worked Exercise 3.18

Prove that the distance property (M2) holds for $d^{(2)}$.

Solution

This follows from the fact that, for any $p, q \in \mathbb{R}$, $(p-q)^2 = (q-p)^2$.

Let $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$. Then

$$\begin{aligned} d^{(2)}(\mathbf{a}, \mathbf{b}) &= \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2} \\ &= \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2} = d^{(2)}(\mathbf{b}, \mathbf{a}). \end{aligned}$$

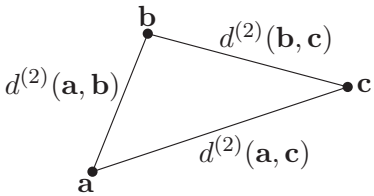


Figure 3.12 The Triangle Inequality

The third important property is the **Triangle Inequality**. This inequality expresses the fact that in a triangle, the length of any one side cannot exceed the sum of the lengths of the other two sides. It can be expressed symbolically as follows, and is illustrated in Figure 3.12.

(M3) For each $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^2$, we have $d^{(2)}(\mathbf{a}, \mathbf{c}) \leq d^{(2)}(\mathbf{a}, \mathbf{b}) + d^{(2)}(\mathbf{b}, \mathbf{c})$.

Exercise 3.7

Verify the Triangle Inequality for the points

$$\mathbf{a} = (3, 4), \quad \mathbf{b} = (6, 0), \quad \mathbf{c} = (5, -3).$$

There are several ways to prove the Triangle Inequality for $d^{(2)}$. We will prove a general version for Euclidean distance on $\mathbb{R}^n$ in the next chapter, obtaining the Triangle Inequality for the plane as the special case $n = 2$.

The three properties (M1)–(M3) also apply to the Euclidean distance function $|b - a|$ on $\mathbb{R}$, for which they were demonstrated in Subsection 1.1. They can be summarised as follows.

(M1) For each $a, b \in \mathbb{R}$, $|b - a| \geq 0$, with $|b - a| = 0$ if, and only if, $b = a$.

(M2) For each $a, b \in \mathbb{R}$, $|b - a| = |a - b|$.

(M3) For each $a, b, c \in \mathbb{R}$, $|c - a| \leq |b - a| + |c - b|$.

The properties (M1)–(M3) hold not only for the Euclidean distance functions defined on $\mathbb{R}$ and $\mathbb{R}^2$, but also for Euclidean distance in higher dimensions. In fact, these three properties are the key properties that we expect any reasonable notion of 'distance' to possess. In Chapter 14, we will specify that *any* set endowed with a function satisfying properties (M1)–(M3) is a set for which distance is defined. We will then be in a position to explore related properties such as convergence of sequences and continuity in a wider setting.

Learning outcomes

We list those topics on which we may set assessment questions to test your understanding of this chapter.

After working through this chapter you should be able to:

(a) State the definition of a function.

(b) Use the modulus function in inequalities and obtain the solution sets of inequalities involving the modulus function.

(c) Use the Triangle Inequality and its variants in $\mathbb{R}$.

(d) Explain the meanings of *sequence*, *null sequence*, *convergent sequence*, *divergent sequence* and *subsequence*.

(e) Test a given real sequence for convergence or divergence.

(f) Use the (sequential) definition of continuity to determine whether a given function from $\mathbb{R}$ to $\mathbb{R}$ is continuous.

(g) Use the rules for continuous functions to determine whether a given function from $\mathbb{R}$ to $\mathbb{R}$ is continuous.

(h) State the Intermediate Value Theorem and use it to show that certain equations have roots.

(i) State the Boundedness and Extreme Value Theorems.

(j) Calculate the distance between any two points in the plane $\mathbb{R}^2$.

(k) Test a given sequence in $\mathbb{R}^2$ for convergence or divergence.

(l) Explain the definition of continuity for functions from $\mathbb{R}^2$ to $\mathbb{R}$.

(m) Determine whether a given function from $\mathbb{R}^2$ to $\mathbb{R}$ is continuous.

(n) Understand the Euclidean distance function on $\mathbb{R}^2$, and appreciate its properties.

Solutions to exercises

Solution to Exercise 1.1

(a) This is a function from $\mathbb{Z}$ to $\mathbb{N}$, since the square of an integer is a non-negative integer and so adding 1 to this number always gives a natural number.

(b) This is not a function from $\mathbb{Z}$ to $\mathbb{N}$, since it associates with a given integer a *set* of natural numbers, and this is not an element of the codomain.

(c) This is not a function from $\mathbb{Z}$ to $\mathbb{N}$, since the integer 0 has no well-defined image. In particular, each non-zero integer n is a factor of 0 and so the set of factors of 0 is $\mathbb{Z} - \{0\}$, which is an infinite set.

Solution to Exercise 1.2

Only the rule given in Exercise 1.1(a) defines a function, $f\colon \mathbb{Z} \to \mathbb{N}$ given by $f(m) = m^2 + 1$. For this function,

$$f(\mathbb{N}) = \{f(n) : n \in \mathbb{N}\} = \{n^2 + 1 : n \in \mathbb{N}\} = \{2, 5, 10, 17, 26, \ldots\}$$

and

$$f^{-1}(\{5\}) = \{x \in \mathbb{Z} : f(x) \in \{5\}\} = \{x \in \mathbb{Z} : x^2 + 1 \in \{5\}\} = \{-2, 2\}.$$

Solution to Exercise 1.3

The answers to parts (a) and (b) can be read off immediately from Proposition 1.2.

(a) The solution set is $(-3, 3)$.

(b) The solution set is $\emptyset$.

(c) $|3x| < 3$ if, and only if, $|3|\,|x| < 3$, which is true if, and only if, $3|x| < 3$, which holds if, and only if, $|x| < 1$. Thus $\{x \in \mathbb{R} : |3x| < 3\} = \{x \in \mathbb{R} : |x| < 1\}$, so the solution set is the interval $(-1, 1)$.

Solution to Exercise 1.4

On noting that $|a| = ||a||$, we can read directly from Worked Exercise 1.3 that

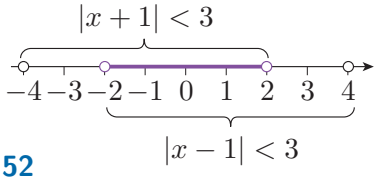
$$|a| = ||a|| \leq |b| \text{ if, and only if, } -|b| \leq |a| \leq |b|,$$

as required.

Solution to Exercise 1.5

(a) By the first result in Proposition 1.5, with $b = 1$ and $c = 3$, the solution set is the open interval $(1 - 3, 1 + 3) = (-2, 4)$.

(b) From part (a), the solution set of $|x - 1| < 3$ is the open interval $(-2, 4)$. Similarly, the solution set of $|x + 1| < 3$ is the open interval $(-4, 2)$. Since both inequalities must hold, we take the intersection of the intervals $(-2, 4)$ and $(-4, 2)$. This is the open interval $(-2, 2)$, as shown.



Solution to Exercise 1.6

We prove the contrapositive: that is, instead of showing that for non-negative numbers $p^2 \leq q^2 \implies p \leq q$, we prove the equivalent statement that for non-negative numbers p and q, if $q < p$, then $q^2 < p^2$.

Suppose that $0 \leq q < p$. Then, since inequalities are preserved when multiplying by a positive real number, $0 \leq q^2 < pq < p^2$. The result follows.

Solution to Exercise 1.7

(a) We use the Triangle Inequality twice.

Let $d = b + c$. Then

$$|a + b + c| = |a + d| \leq |a| + |d|,$$

by the Triangle Inequality. Also,

$$|d| = |b + c| \leq |b| + |c|.$$

Combining these two inequalities gives

$$|a + b + c| \leq |a| + |b| + |c|.$$

(b) The Triangle Inequality states that, for *all* real numbers x and y,

$$|x + y| \leq |x| + |y|.$$

Let $x = a$ and $y = -b$. Then

$$|a - b| = |a + (-b)| \leq |a| + |-b| = |a| + |b|.$$

Solution to Exercise 1.8

For real numbers a and b, we use the Triangle Inequality to find

$$|b| = |(b - a) + a| \leq |b - a| + |a|$$

and

$$|a| = |(a - b) + b| \leq |a - b| + |b| = |b - a| + |b|.$$

Hence

$$|b| - |a| \leq |b - a|$$

and

$$|a| - |b| \leq |b - a|.$$

In particular, since $\big||b| - |a|\big|$ equals one of $|b| - |a|$ or $|a| - |b|$,

$$\big||b| - |a|\big| \leq |b - a|,$$

as required.

Solution to Exercise 1.9

Let $\varepsilon > 0$ be given. Here $|a_n| = \frac{1}{\sqrt{n}}$ and

$$\frac{1}{\sqrt{n}} < \varepsilon \quad \text{if, and only if,} \quad n > \frac{1}{\varepsilon^2}.$$

If we set $N = \text{int}\left(\frac{1}{\varepsilon^2}\right)$ then

$$|a_n| = \frac{1}{\sqrt{n}} < \varepsilon \quad \text{whenever } n > N.$$

Hence (a_n) is null.

Solution to Exercise 1.10

For each n, $|a_n| = |2| = 2$. We take any $\varepsilon \leq 2$, say $\varepsilon = 1$, and so, for each n, $|a_n| = 2 > 1 = \varepsilon$. We deduce that (a_n) is not a null sequence.

Solution to Exercise 1.11

Since (a_n) is null, the Multiple Rule implies that (λa_n) is null. Now $|b_n| \leq \lambda |a_n| = |\lambda a_n|$ (since $\lambda \geq 0$). It follows from the Squeeze Rule that (b_n) is null.

Solution to Exercise 1.12

A sequence (a_n) is divergent if for each real number l, the sequence $(a_n - l)$ is not a null sequence. That is, for each $l \in \mathbb{R}$, there is at least one $\varepsilon > 0$ for which there is *no* natural number N such that $|a_n - l| < \varepsilon$ whenever $n > N$. Alternatively, whatever the value of l, there exists an $\varepsilon > 0$ such that, no matter what choice of N we make, we can always find a natural number $n > N$ for which $|a_n - l| \geq \varepsilon$.

Solution to Exercise 1.13

Let

$$a_n = \frac{4n^2 + 1000^n}{n! + 7n^3}.$$

Dividing both the numerator and denominator of a_n by the dominant term $n!$ gives

$$a_n = \frac{4(n^2/n!) + 1000^n/n!}{1 + 7(n^3/n!)}.$$

The sequences $\left(\frac{n^2}{n!}\right)$, $\left(\frac{1000^n}{n!}\right)$, $\left(\frac{n^3}{n!}\right)$ are all basic null sequences and any constant sequence converges to its constant value. Hence, by the combination rules for convergent sequences,

$$a_n \to \frac{(4 \times 0) + 0}{1 + (7 \times 0)} = 0 \quad \text{as } n \to \infty.$$

Thus (a_n) converges to 0, i.e. it is null.

Solution to Exercise 1.14

Write $a_n = 2^n$ and let $n_k = k^2$ for each $k \in \mathbb{N}$. Then (n_k) is a strictly increasing sequence of positive integers, and $a_{n_k} = 2^{k^2}$. Thus $(a_{n_k}) = (2^{k^2})$ is a subsequence of $(a_n) = (2^n)$.

Solution to Exercise 1.15

Let

$$a_n = (-1)^n \frac{1-n}{n}.$$

Consider the even and odd subsequences (a_{2k}) and (a_{2k-1}).

For (a_{2k}), we have

$$a_{2k} = (-1)^{2k} \frac{1-2k}{2k} = \frac{1-2k}{2k}.$$

Dividing both the numerator and denominator by the dominant term k gives

$$a_{2k} = \frac{\frac{1}{k} - 2}{2}.$$

Now $(\frac{1}{k})$ is a basic null sequence, so the combination rules for convergent sequences imply that

$$a_{2k} \to \frac{0-2}{2} = -1 \quad \text{as } k \to \infty.$$

For (a_{2k-1}), we have

$$\begin{aligned} a_{2k-1} &= (-1)^{2k-1} \frac{1-(2k-1)}{2k-1} \\ &= (-1)\frac{2-2k}{2k-1} = \frac{2k-2}{2k-1}. \end{aligned}$$

Dividing both the numerator and denominator by the dominant term k gives

$$a_{2k-1} = \frac{2 - \frac{2}{k}}{2 - \frac{1}{k}}.$$

Now $(\frac{1}{k})$ is a basic null sequence and so, by the Multiple Rule for null sequences (with $\lambda = 2$), $(\frac{2}{k})$ is a null sequence. Hence the combination rules for convergent sequences imply that

$$a_{2k-1} \to \frac{2-0}{2-0} = 1 \quad \text{as} \quad k \to \infty.$$

Since the even and odd subsequences have different limits, we deduce from Proposition 1.25 that (a_n) is divergent.

Solution to Exercise 1.16

(a) This is a convergent sequence. To prove this, we use the combination rules for convergent sequences together with our list of basic null sequences.

The sequences $\left(\frac{n^3}{n!}\right)$ and $\left(\frac{1}{n^2}\right)$ are basic null sequences. Hence by the Multiple Rule, so are $\left(\frac{3n^3}{n!}\right)$ and $\left(\frac{7}{n^2}\right)$, and so, by the Sum Rule, so is $\left(\frac{3n^3}{n!}+\frac{7}{n^2}\right)$. Thus, the limit is 0.

(b) This sequence does not converge. We have

$$(n+1)^2 - n^2 = 2n+1.$$

Thus $((n+1)^2 - n^2) = (2n+1)$, which is divergent. To see that $(2n+1)$ is divergent, we show directly that if $a \in \mathbb{R}$, then $2n+1 \not\to a$ as $n \to \infty$.

If $\varepsilon = 1$ and $N \in \mathbb{N}$ is chosen so that $N \geq \frac{|a|}{2}$, then for $n > N$,

$$\begin{aligned}|(2n+1) - a| &\geq 2n + 1 - |a| > 2N + 1 - |a| \\ &\geq 1 = \varepsilon.\end{aligned}$$

(We have used the Reverse Triangle Inequality here.) We conclude that $(2n+1)$ does not converge to a, and so, as a is arbitrary, does not converge to anything.

(c) This is a convergent sequence. We first observe that, for each $n \in \mathbb{N}$,

This is a 'trick' that is worth knowing.

$$\begin{aligned}\sqrt{n+1} - \sqrt{n} &= \frac{(n+1)-n}{\sqrt{n+1}+\sqrt{n}} \\ &= \frac{1}{\sqrt{n+1}+\sqrt{n}}.\end{aligned}$$

Hence

$$|\sqrt{n+1} - \sqrt{n}| = \left|\frac{1}{\sqrt{n+1}+\sqrt{n}}\right| \leq \frac{1}{\sqrt{n}},$$

which is a basic null sequence. We deduce from the Squeeze Rule that $(\sqrt{n+1} - \sqrt{n})$ is null.

Solution to Exercise 2.1

Let a be a general point in $\mathbb{R}$. We show that if (x_n) is a sequence in $\mathbb{R}$ that converges to a, then the sequence $(f(x_n))$ converges to $f(a) = a^3 + 3$.

So let (x_n) be a sequence (in $\mathbb{R}$) that converges to a. Then

$$f(x_n) = x_n^3 + 3 = (x_n \times x_n \times x_n) + 3.$$

Thus, since constant sequences converge to their constant value, we apply the Product Rule for convergent sequences (twice) together with the Sum Rule for convergent sequences and deduce that

$$f(x_n) \to (a \times a \times a) + 3 = a^3 + 3 = f(a) \text{ as } n \to \infty.$$

Hence, $f(x_n) \to f(a)$ as $x_n \to a$, and so f is continuous at a.

Since a is an arbitrary point of $\mathbb{R}$, we conclude that f is continuous on $\mathbb{R}$.

Solution to Exercise 2.2

Let $a \neq 0$. There are two possibilities: either $a > 0$ or $a < 0$.

If $a > 0$, then $f_2(a) = 1$. Suppose that (x_n) is a sequence in $\mathbb{R}$ for which $x_n \to a$ as $n \to \infty$.

Then we can find $N \in \mathbb{N}$ such that $x_n > 0$ for $n > N$ (for example by taking N so that, for $n > N$, $|x_n - a| < |a|$ – ensuring the distance of x_n from a is less than the distance of a from 0).

Thus for $n > N$, $f_2(x_n) = f(a) = 1$, so $f_2(x_n) \to 1 = f_2(a)$ as $n \to \infty$. Hence f_2 is continuous at a.

For $a < 0$, we proceed in a similar manner. Upon noting that $f_2(a) = 0$, we suppose (x_n) is a sequence in $\mathbb{R}$ that converges to a.

Then we can find $N \in \mathbb{N}$ such that $x_n < 0$ for $n > N$, so that $f_2(x_n) = f_2(a) = 0$ for $n > N$. Hence $f_2(x_n) \to 0 = f_2(a)$ as $n \to \infty$, and f_2 is continuous at a.

Since we have covered all possible values of $a \neq 0$, we conclude that f is continuous on $\mathbb{R} - \{0\}$.

Solution to Exercise 2.3

We need to show that for any *null* sequence (x_n), $f(x_n) \to f(0) = 0$.

So suppose (x_n) is null. Then, for each n, *either* $x_n \neq 0$ and $|f(x_n)| = |x_n \cos(\pi/x_n)| = |x_n|\,|\cos(\pi/x_n)|$, *or* $x_n = 0$ and $f(x_n) = 0$.

Now $|\cos a| \leq 1$ for all real numbers a and so, when $x_n \neq 0$,

$$|f(x_n)| = |x_n|\,|\cos(\pi/x_n)| \leq |x_n|.$$

Whereas if $x_n = 0$, then

$$|f(x_n)| = 0 \leq |x_n| = 0.$$

Hence $|f(x_n)| \leq |x_n|$ for each $n \in \mathbb{N}$.

But (x_n) is a null sequence and so, by the Squeeze Rule, $|f(x_n)| \to 0$ as $n \to \infty$. Thus $f(x_n) \to f(0) = 0$ as $x_n \to 0$.

Hence f is continuous at 0.

Solution to Exercise 2.4

(a) For $x \neq 1$,

$$f(x) = \frac{x^2 - 1}{x - 1} = \frac{(x+1)(x-1)}{x-1} = x + 1$$

and so if (x_n) is a sequence in $\mathbb{R} - \{1\}$, then

$$f(x_n) = x_n + 1.$$

So, if $x_n \to 1$ as $n \to \infty$, then the Sum Rule implies that

$$f(x_n) = x_n + 1 \to 1 + 1 = 2 \text{ as } n \to \infty.$$

Hence $(f(x_n))$ is a convergent sequence with limit 2.

(b) To show that g is continuous at $a = 1$, suppose that $x_n \to 1$ as $n \to \infty$. Then for any given n, either $x_n = 1$ and so $g(x_n) = 2$, or $x_n \neq 1$ and $g(x_n) = f(x_n) = x_n + 1$. In either case,

$$|g(x_n) - g(1)| = |g(x_n) - 2| \leq |x_n - 1| \to 0 \quad \text{as } n \to \infty.$$

We deduce that $g(x_n) \to g(1)$, and hence g is continuous at $a = 1$.

In fact, for $x \neq 1$, $f(x) = x + 1$ and so it is no surprise that g is continuous at 1.

Solution to Exercise 2.5

The functions $x \mapsto e^x$, $x \mapsto -4x^3$ and $x \mapsto \sin x$ are basic continuous functions with domain $\mathbb{R}$. The function $x \mapsto 1/x$ is a basic continuous function with domain $\mathbb{R} - \{0\}$. Hence, by the Restriction Rule, $x \mapsto 1/x$ and $x \mapsto -4x^3$ are continuous on $(0, \infty)$. Also, by the Composition Rule, $x \mapsto \sin(1/x)$ and $x \mapsto e^{-4x^3}$ are continuous on $(0, \infty)$. It follows from the Multiple Rule that $x \mapsto 2\sin(1/x)$ and $x \mapsto -3e^{-4x^3}$ are continuous on $(0, \infty)$. Finally, by the Sum Rule, f is continuous on $(0, \infty)$.

Solution to Exercise 2.6

We are given that f is continuous on $\mathbb{R}$. The function $x \mapsto |x|$ is a basic continuous function on $\mathbb{R}$. Hence, by the Composition Rule, the function $|f|$ is continuous on $\mathbb{R}$.

Solution to Exercise 2.7

We first show that $g\colon [0,1] \to \mathbb{R}$ given by $g(x) = f(x) - x$ is continuous. Since $x \mapsto -x$ is a basic continuous function on $\mathbb{R}$, it is continuous on $[0,1]$ (by the Restriction Rule). Since f is continuous on $[0,1]$, so is $g(x)$ (by the Sum Rule).

Now notice that, since $f(0) \geq 0$,

$$g(0) = f(0) - 0 \geq 0$$

and, since $f(1) \leq 1$,

$$g(1) = f(1) - 1 \leq 1 - 1 \leq 0.$$

Hence *either* $g(0) = g(1) = 0$, implying that $f(0) = 0$ and $f(1) = 1$ (and so 0 and 1 are both fixed points of f), *or* $g(0) > g(1)$ and we can use the Intermediate Value Theorem to deduce that there is a number $k \in [0,1]$ for which $g(k) = 0$. But if $g(k) = 0$, then $f(k) - k = 0$, and so $f(k) = k$.

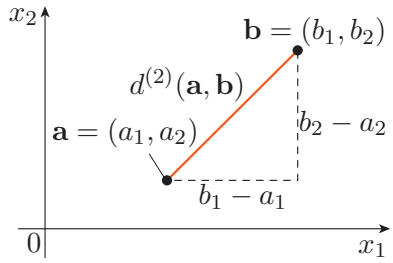
Solution to Exercise 2.8

The function f is the restriction of a basic continuous function, and so is continuous on $[-1,1] - \{0\}$. However, the set $[-1,1] - \{0\}$ is *not* a closed interval, and so the hypotheses of the Boundedness Theorem and the Extreme Value Theorem are not satisfied. So we cannot expect the conclusions of the Boundedness and Extreme Value Theorems to hold – and indeed, they do not: f is bounded neither above nor below, and so certainly cannot attain its bounds.

Solution to Exercise 3.1

Let $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$, as shown. Then

$$\mathbf{a} - \mathbf{b} = (a_1 - b_1, a_2 - b_2),$$
$$\mathbf{b} - \mathbf{a} = (b_1 - a_1, b_2 - a_2).$$



Thus,

$$\begin{aligned} &d^{(2)}(\mathbf{a} - \mathbf{b}, \mathbf{0}) \\ &\quad = \sqrt{(0 - (a_1 - b_1))^2 + (0 - (a_2 - b_2))^2} \\ &\quad = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}. \end{aligned}$$

Similarly,

$$\begin{aligned} &d^{(2)}(\mathbf{b} - \mathbf{a}, \mathbf{0}) \\ &\quad = \sqrt{(0 - (b_1 - a_1))^2 + (0 - (b_2 - a_2))^2} \\ &\quad = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}. \end{aligned}$$

Also

$$d^{(2)}(\mathbf{a}, \mathbf{b}) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}.$$

Thus, $d^{(2)}(\mathbf{a}, \mathbf{b}) = d^{(2)}(\mathbf{a} - \mathbf{b}, \mathbf{0}) = d^{(2)}(\mathbf{b} - \mathbf{a}, \mathbf{0})$.

Solution to Exercise 3.2

The terms of $(\mathbf{b}_n)$ appear to get near to $(1,4)$ as n gets large, and so we show that $(d^{(2)}(\mathbf{b}_n,(1,4))^2)$ is a real null sequence. For $n \in \mathbb{N}$, we have

$$\begin{aligned} d^{(2)}\left(\left(\tfrac{3}{n}+1, 4-\tfrac{2}{n^3}\right),(1,4)\right)^2 &= \left(\tfrac{3}{n}\right)^2 + \left(-\tfrac{2}{n^3}\right)^2 \\ &= \tfrac{9}{n^2} + \tfrac{4}{n^6}, \end{aligned}$$

a sum of multiples of basic null sequences. Hence $(d^{(2)}(\mathbf{b}_n,(1,4))^2)$ is a null sequence.

We conclude that $(\mathbf{b}_n)$ converges to the limit $(1,4)$.

Solution to Exercise 3.3

(a) By Theorem 3.7, it is enough to show that the coordinate sequences $\left(\dfrac{n^2}{3^n}\right)$ and $\left(\dfrac{3n!+2^n}{n^{50}+n!}\right)$ both converge. We note that $\left(\dfrac{n^2}{3^n}\right)$ is a basic null sequence, while in Worked Exercise 1.20(b) we showed that $\left(\dfrac{3n!+2^n}{n^{50}+n!}\right)$ converges to 3.

Therefore, using Theorem 3.7, we find that $\mathbf{a}_n \to (0,3)$ as $n \to \infty$.

(b) Since $\mathbf{a}_n \to \boldsymbol{\ell}$ and $\mathbf{b}_n \to \mathbf{m}$ as $n \to \infty$, it follows from Theorem 3.7 that

$$a_{1,n} \to l_1,\ a_{2,n} \to l_2,\ b_{1,n} \to m_1 \text{ and } b_{2,n} \to m_2 \text{ as } n \to \infty.$$

Hence, by the Sum Rule for real sequences,

$$c_{1,n} = a_{1,n} + b_{1,n} \to l_1 + m_1$$

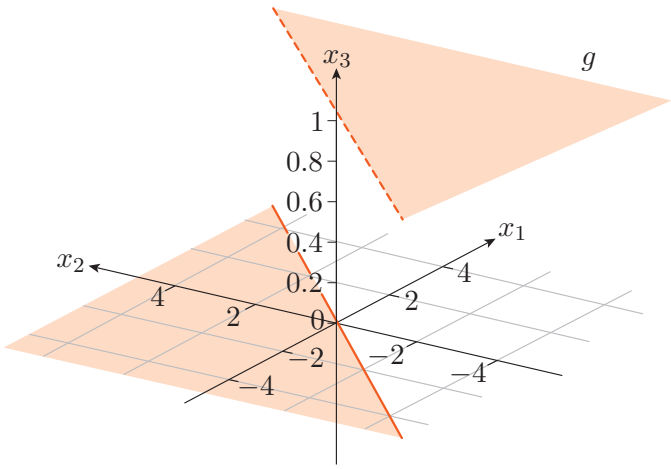
and

$$c_{2,n} = a_{2,n} + b_{2,n} \to l_2 + m_2$$

as $n \to \infty$. We now use Theorem 3.7 again to deduce that $\mathbf{c}_n \to \boldsymbol{\ell} + \mathbf{m}$ as $n \to \infty$.

Solution to Exercise 3.4

The function g (shown in the figure) appears to be discontinuous along the line $x_1 = x_2$: the function values on one side of the line are all equal to 1, and on the other side (and on the line itself) are equal to 0. There are points (x_1, x_2) arbitrarily close to $(2, 2)$ for which $g(x_1, x_2) = 0$, and other points for which $g(x_1, x_2) = 1$.



To prove that g is discontinuous at $(2, 2)$, we need only to find one sequence $\mathbf{x}_n \to (2, 2)$ such that $g(\mathbf{x}_n) \not\to g(2, 2) = 0$.

To do this we choose a sequence that approaches $(2, 2)$ from the $x_1 > x_2$ side, such as $\mathbf{x}_n = (2 + \frac{1}{n}, 2 - \frac{1}{n})$ for $n \in \mathbb{N}$. Then

$$d^{(2)}(\mathbf{x}_n, (2, 2)) = \frac{\sqrt{2}}{n} \to 0 \text{ as } n \to \infty,$$

and so $\mathbf{x}_n \to (2, 2)$ as $n \to \infty$. However,

$$g(\mathbf{x}_n) = g(2 + \tfrac{1}{n}, 2 - \tfrac{1}{n}) = 1, \quad \text{for each } n \in \mathbb{N},$$

and so the sequence $(g(\mathbf{x}_n))$ does not converge to $g(2, 2) = 0$ as $n \to \infty$.

Hence g is discontinuous at $(2, 2)$.

Solution to Exercise 3.5

The function f consists of adding the variables x_1 and x_2 and then taking the sine of the result. The variables x_1 and x_2 are given by the projection functions $p_1(\mathbf{x}) = x_1$ and $p_2(\mathbf{x}) = x_2$, both of which are continuous. Hence, by the Sum Rule, so is $p_1(\mathbf{x}) + p_2(\mathbf{x}) = x_1 + x_2$. This function has image set $\mathbb{R}$, which is the domain of the sine function. The sine function is a basic continuous function. Therefore, by the Composition Rule, we can compose it with $p_1(\mathbf{x}) + p_2(\mathbf{x})$ to give the continuous function $\sin(p_1(\mathbf{x}) + p_2(\mathbf{x})) = \sin(x_1 + x_2) = f(\mathbf{x})$. Thus f is continuous on $\mathbb{R}^2$.

Solution to Exercise 3.6

The variables x_1 and x_2 are given by the projection functions p_1 and p_2, both of which are continuous. Hence so is $-p_1(\mathbf{x})p_2(\mathbf{x}) = -x_1x_2$, by the Multiple and Product Rules. This function has image set $\mathbb{R}$, which is the domain of the function exp. Since exp is a basic continuous function, so is $\exp(-x_1x_2) = e^{-x_1x_2} = f(\mathbf{x})$, by the Composition Rule. Thus f is continuous on $\mathbb{R}^2$.

Solution to Exercise 3.7

$$
\begin{aligned}
&d^{(2)}(\mathbf{a},\mathbf{b}) + d^{(2)}(\mathbf{b},\mathbf{c}) \\
&\quad = \sqrt{(6-3)^2 + (0-4)^2} + \sqrt{(5-6)^2 + (-3-0)^2} \\
&\quad = 5 + \sqrt{10} > 5 + \sqrt{9} = 8;
\end{aligned}
$$

$$
d^{(2)}(\mathbf{a},\mathbf{c}) = \sqrt{(5-3)^2 + (-3-4)^2} = \sqrt{53} < \sqrt{64} = 8.
$$

Hence, $d^{(2)}(\mathbf{a},\mathbf{c}) \leq d^{(2)}(\mathbf{a},\mathbf{b}) + d^{(2)}(\mathbf{b},\mathbf{c})$.